

Geneva meeting the best expected

Only the most unrestrained optimists expected written agreement last week between the United States and the Soviet Union on arms control. It is sufficient that there is to be another meeting.

UNREASONABLE optimism, far from being reprehensible, is more often one of the engines of beneficent change. How else could the New World, encompassing the United States, have been discovered? But unreasonable optimism is an encumbrance when its consequences include unreasonable disappointment, of which there has been far too much in the week following the meeting at Geneva between President Ronald Reagan and Mr Mikhail Gorbachev, general secretary of the Communist Party of the Soviet Union. The fact is that, since high summer, there has not been the slightest chance that the two men in person, however friendly they chose to be, would cut through the difficulties that have defeated their diplomats over several months. The chance that Geneva would bring a specific agreement on strategic arms control had shrunk to nothing by mid-September, and for reasons that had become generally understood. The best that could be hoped for is that the two meeting at Geneva would agree to meet again (see *Nature* 14 November, p.91). In reality, they have done twice as well as that, and have agreed to meet again in 1986 and 1987. It would have been a bad business for all of us if this had not come about. So it is reprehensible that so many newspapers and other sources of public comment should be saying that nothing happened at Geneva. That the two sides did not fall out is what happened.

Where things go from here is the absorbing question. The statements at the end of the Geneva meeting are fuller and emptier than might have been expected. The two superpowers have gone out of their way, with the review conference of the Non-Proliferation Treaty successfully just behind them, to reaffirm their opposition to the spread of nuclear weapons, but have not explicitly said that they will honour, let alone strengthen, the Anti-Ballistic Missile Treaty, the properly ratified agreement on arms control threatened by some of the statements in favour of anti-ballistic missile defences put out from the United States in recent months. Similarly, nothing was said about what will happen when the SALT II treaty (unratified) expires at the end of the year; presumably the calculation is that this question is subsumed in the declaration that the permanent negotiators on arms control at Geneva should accelerate their search for an agreement on strategic arms. Meanwhile, the Soviet side makes no secret of its distaste for US plans for the Strategic Defense Initiative, while the United States can claim some credit for having being allowed to stand firm.

What, then, will happen next? The most urgent task to be tackled once the negotiators reassemble at Geneva is that they should trim their sails to what their leaders say they have agreed. The goal seems to be a 50 per cent reduction of strategic arms, but it is not clear whether the count is to be made by launchers or by warheads. The difference is easily bridgeable, but could also easily be magnified into an insuperable obstacle. The two leaders, having made their journey to Geneva and having claimed success, have a continuing responsibility to ensure not merely that their negotiators are suitably flexible but that the people not privy to the negotiations, the rest of us, are reassured from time to time by public rehearsals of what the understanding is supposed to have been. Otherwise, especially in openly warring Washington, there is a risk that the objective will be corrupted or misunderstood.

Meanwhile, there is a strong case for paying more explicit

attention to the softer parts of the Geneva agreement, those in which the participants promised fuller cooperation on a variety of issues, cultural and scientific exchanges prominent among them. Such aspirations are commonplace when statesmen meet and have little else to agree about. The backroom boys can afterwards be asked to say what scientific collaboration means, and can draw up more detailed agreements for signature at later meetings. The process is almost a formality. Grave officials of learned academies settle on new numbers to describe what numbers of man-months of reciprocal visits shall be allowed on either side. Lists of priorities are drawn up specifying what each side would like to learn from the other, but with the reservation that nothing sensitive in the military sense should be included. Then, when everything is signed and sealed, it turns out that one side or the other is unable or unwilling to make the agreement work. In the past decade or so, for example, it has usually been impossible for Western academies to fill their quota of man-months allowed in the Soviet Union. It is a step in the right direction that, on this occasion, the United States and the Soviet Union have specified, vaguely it must be acknowledged, thermonuclear fusion as a field in which collaboration might be worthwhile. But, yet again, they have been putting the cart before the horse.

The reality is that the formal agreements reached between academies of countries otherwise hostile to each other are never satisfactory. What is needed in the dealings between the United States and the Soviet Union on scientific collaboration is not a set of quotas and a list of priorities but a mechanism by which scholars in both places can decide for themselves what kinds of collaborative research they would like to undertake. Thus the Soviet Union should (and probably does) know that there are many people in the West who would give their eye teeth to work at some Soviet archaeological site, or have a chance to tackle some geological investigation in the Soviet Union. It goes without saying that there are different fields of science in the United States in which Soviet scientists are eager to participate. The obvious difficulties — that both sides are sensitive about security — are well understood, and could be more explicitly defined. Why not, within such a framework, agree that there should be a pot of money to facilitate exchanges and then agree that the projects on which it is spent should be determined by those directly involved, not by their governments' civil servants? To follow past patterns is to invite failure and, ultimately, to bring science into disrepute unfairly. □

Unity by common fright

Europe's plan to collaborate in high technology seems to have struck a welcoming chord.

EUROPE seems to have made an unexpected discovery about itself. When the government of France issued its rallying call to the rest of Europe that there should be a collaborative programme of technological research and development, it was natural that other governments in Western Europe should be suspicious of the motives. For one thing, France did not hide its belief that such a programme was a necessary counterweight to the US Strategic Defense Initiative (SDI), which cannot but be a power-



ful means of stimulating high technology in the United States; the snag was that many governments had good reasons for not opting out of SDI. But, equally, there was no way of making sure that the plan called Eureka was not another way of advancing French claims to European leadership in technology, or perhaps even a way of persuading other governments to contribute to French advancement. For why else, the sceptics were asking earlier in the year, has France been urging technical collaboration on the rest of Europe at least since 1968, when M. Pierre Aigrain (then in the French government, now a director of the French Thomson company) tried on his government's behalf to persuade the European Communities to joint action? During M. François Mitterrand's presidency, the exhortations to technical collaboration have been even more frequent and insistent.

The surprise is that other governments in Western Europe seem at last to have taken to listening to what the French are saying. After two decades of scepticism, the belief that France may have been right all along has begun to gain ground. So much, at least, appears to have been the impression they created at the ministerial meeting on Eureka held at Hannover earlier this month. What seems to have happened in this economically disappointing year is that Western Europe has become collectively despondent about its technical future. Most governments are trying hard to succeed in high technology; most of them are presented every other day with evidence of how difficult it is to succeed in the face of competition from Japan and the United States. It is not surprising that one participant at Hannover said that European governments are now "running scared".

So, almost by chance, Mitterrand's Eureka will succeed when his and his predecessors' efforts in the same direction have mostly failed. The French president deserves credit for his persistence, the other governments of Western Europe for their new-found realism. But it is one thing to be agreed on a common cause and quite another to be sure of attaining it. The practical question still to be answered is whether mechanisms can be found that will enable Eureka to deliver the prizes its supporters promise. Two features of Eureka have so far been agreed. The framework of collaboration will be wider than the European Communities as such, so as to include Austria, Sweden and Switzerland in particular. And the programme will be supported not by another huge pot of public money, but by the funds of those most likely to benefit, the companies that decide to engage in transnational projects with each other.

On the face of things, it is not obvious why European companies should not collaborate with each other on worthwhile technical projects without France's banner labelled Eureka. Although the European Communities have regulations to prevent anti-competitive collusion by competing companies, these are neither as explicit as in the United States, nor are the penalties of transgression as fearsome. Part of the explanation for the poverty of international collaboration in Europe in past decades must surely be the conservatism of even progressive European industrial companies, many of whom still consider themselves as bitter rivals for the same block of overseas trade, that with the United States in particular. There is a chance that exhortation by itself may change this state of affairs, in which case Eureka may be very cheap. More probably, there are technical projects yet to be devised where public money will help to oil rusty wheels.

European governments should also recognize that the kind of collaboration foreseen by Eureka cannot be the basis of long-term economic success. It may help, in the short run, if more companies collaborate on the development of improved products, but either the agreements will then lapse and the companies will revert to rivalry, or they will find they have a successful product which they can market successfully only through some commonly owned channel, perhaps even an independent company in its own right. They will also discover, on the way to success of that kind, that the most wholehearted collaboration cannot succeed if the products whose development is intended are not sold competitively. The lesson from Japan on this point is clear; Japanese companies have succeeded in the export of

high-quality technological products only because their domestic market is even more demanding than those abroad. So will European governments follow the logic of the enterprise on which they have engaged in Eureka, and agree that success entails the banishment of restrictive practices on public purchasing; of the regulations and unwritten rules by which governments in Europe try to preserve the national identity of their domiciled companies; and of the barriers that persist to the free movement of academics and their students from one part of Europe to another? Unless governments agree that these impediments to efficiency are long-term barriers to success in high technology, Eureka will fail. □

Honest pennies to earn

Britain's environment research council should count against the costs of success.

THE British Natural Environment Research Council may fear that there is no way in which it can win friends, or placate its critics, but there is an awkward contradiction built into the way it is compelled to operate on a shrinking public budget. The framework of the inconsistency dates to the Rothschild reorganization of British civil science in 1971, but the issue has been sharpened only in the past five years. Rothschild, it will be remembered, argued that part of publicly supported research should be regarded as applied research, which should by rights be paid for by the ultimate beneficiaries (called customers), whose proxies would be government departments. So, when the British government accepted the formula, nearly a half of the council's budget (which proportion has since shrunk), was transferred to departments such as agriculture (interested in environmental consequences of fertilizer usage), the environment (interested in the environment) and energy (interested in natural resources). With the passage of time, departmental interests have been attenuated, and their willingness to pay for research has correspondingly declined. So the research council, to keep its establishments occupied, has been forced into the market, selling itself as a research contractor.

Earlier this week (see p.307), the council was rightly boasting of its success in selling its technique of side-scanning sonar to the US Geological Survey, which has commissioned it to carry out a survey of the continental shelf off the United States. Now the technique has been licensed to a British company which will, no doubt, be able to establish a profitable business elsewhere overseas, from which the council will also (rightly) profit. But there are other fields in which the council's success as a contractor for research has not been as welcome. Thus the British Geological Survey, originally established so as to provide the United Kingdom with a knowledge of the rocks and resources on which it rests, has been forced by penury to set up as a contractor for a wide variety of geological survey work, much of it well within the field of competence of private geological survey companies. This is one of the minor bones of contention that will no doubt be dealt with by the committee of inquiry now under way.

The perplexing question is how it can be right, even admirable, that the council should sell its sonar technology commercially while being criticized for competing with private organizations in geological survey work. The superficial answer is that there is a vast difference between a technique (such as side-scan sonar) where there are no immediate competitors and one (such as geological survey) where there are all too many. The more durable answer is that the role of publicly supported research enterprises working in the field of applications should be to stimulate companies working in the private sector, never to compete with them. It is probably a fault of the British economic system that there was no small company eager to build a future around the side-scanning sonar that is now a feather in the council's cap, in which case the council should have been worrying to stimulate a company of that kind. What is needed for the future is a mechanism for telling when public research must stop acting as an entrepreneur, becoming a midwife for others. □

US-Soviet fusion

US still vague about Geneva agreement

Washington

FUSION energy researchers in the United States were optimistic last week that the accord reached in Geneva between President Ronald Reagan and Soviet leader Mikhail Gorbachev on magnetic fusion would lead to increased contacts with their Soviet counterparts. Details of what exactly has been agreed were sparse last week, but the two leaders advocated "the widest possible international cooperation" on development of magnetic fusion power. There are still formidable obstacles, however, in the way of building a joint fusion research reactor.

Fusion research does not rank high

Honour for honoured

THE Technical University in Munich (TUM) is trying to woo back Klaus von Klitzing, winner of this year's Nobel Prize for Physics (see *Nature* 24 October, p.667) by offering him the directorship of a new institute in semiconductor physics. Setting-up costs for the institute are estimated at more than DM20 million, of which it is hoped industry will provide one-third, the other two-thirds coming from the Bavarian state government. Von Klitzing is said to have indicated interest, but is undecided about accepting.

From 1980 to 1984, von Klitzing was an associate professor at TUM but in a reorganization of the physics faculty he was not promoted to a full professorship, so he accepted a directorship at the Max Planck Institute of solid-state physics in Stuttgart. At the new institute, he would be a full professor and will have the opportunity to teach, which he is said to want.

The institute will be set up to provide the most advanced research facilities. It will be divided into three sections, each headed by a director — von Klitzing's colleague Frederik Koch, who proposed the original idea for the institute, has been invited to lead one of those sections. The plan also includes six scientists and nine technicians as well as administrative and secretarial staff. All of these are new positions and have to be included in TUM's next budget. Land is already available but finance for a new building and equipment will be discussed early in December when the state government meets with various industrial concerns.

The institute will be named after the Berlin physicist Walter Schottky. TUM intends to do all it can to lure von Klitzing back, not just because of his recent honour, but because it considers semiconductor physics an important field. Jen Altman

among the issues needing discussion between the superpowers, which perhaps explains the vagueness of the public statements, although further details may yet emerge. It will be for the Department of Energy to draw up detailed proposals for an agenda on collaboration, which could be discussed when Reagan and Gorbachev meet next year in the United States.

There are several reasons why magnetic fusion is a good candidate for international cooperation. Most of the literature is unclassified and the next generation of research machines (which will seek to attain plasma ignition, the point at which fusion becomes self-sustaining) will be enormously expensive.

The US magnetic fusion research budget, though faring well by comparison with other energy research projects, has declined substantially in recent years, from \$464 million in 1984 to \$429 million in 1985, with the budget for fiscal year 1986 down to \$382 million. The programme's goals have correspondingly been pushed into the future. Dr Harold Furth, head of the Princeton Tokamak Fusion Test Reactor, says that because of budget constraints, plans to introduce tritium (the fusion fuel) into the Princeton reactor have been put back until at least the end of 1988. No decisions have been made on a new research machine for the United States in 10 years. Cooperation with the Soviets would be one way to make a limited budget go further.

Soviet magnetic fusion research is comparable with that in the United States. Collaboration takes place at present under an intergovernmental bilateral agreement first negotiated in 1973 and extended in 1983. The number and the standing of scientists in reciprocal exchanges are, however, widely believed to have declined in recent years (partly because of protests over Soviet human rights abuses and its invasion of Afghanistan). There were six scientific teams exchanged last year. Dr Frank Press, president of the National Academy of Sciences (NAS) (whose personal efforts to further individual US-Soviet exchanges have been criticized by anti-Soviet hardliners in the Pentagon) said the Reagan-Gorbachev agreement will affect exchanges "productively".

In the short term, there are likely to be many benefits from increased cooperation on basic plasma physics research and on design and engineering of fusion equipment and diagnostics. But if increasing the frequency of scientific visits is a feasible goal, there remain serious difficulties for cooperation on a fusion reactor. Dr John

Richardson of NAS says that a role can be seen for two machines, one to demonstrate plasma ignition and a second to examine engineering concepts for a future commercial reactor. And there are several ancillary projects that might be tackled as collaborative ventures, such as compact toruses, magnetic mirror fusion and hybrid reactors whose neutron flux would be used to manufacture fuel for conventional fission reactors.

But Richardson can see no "natural division of labour" for a truly international four-way agreement to include Europe, the United States, the Soviet Union and Japan; so far, it has not even been possible to negotiate such a deal with Europe.

Tim Beardsley

Geneva summit

Soviet optimism on fusion plan

THE Soviet response to the prospect that the Geneva summit could lead to increased scientific cooperation has been slower than that in the United States, but is no less enthusiastic. Mr Mikhail Gorbachev himself has been reluctant to "deal with the nuances" of possible cooperation projects, as he phrased it at Geneva last week, preferring to stress the funds that could be released to fight world hunger if the arms race were halted. Indeed, when he did commit himself on a specific project — fusion research — what he had in mind was not bilateral but multilateral cooperation. "It was decided", he said, "jointly to approach a number of other states about cooperation in the field of thermonuclear research. This is a very interesting idea which could guarantee for mankind an inexhaustible source of energy."

The view of the Soviet scientific community was reflected last week in Moscow by Dr Petr Timofeev, deputy director of the Institute of Geology of the Academy of Sciences, who said that the institute had been involved in a number of cooperation projects with the United States on the geology of the oceans which had, regretfully, been curtailed in recent years, but would now, he hoped, be renewed. Other potential fields of cooperation in which Soviet scientists are interested include space research, the "conquest of human disease", the environment (a delegation from the US Environmental Protection Agency visited the Soviet Union just before the Geneva summit) and, of course, controlled thermonuclear fusion.

If this cooperation project does materialize, it could prove a particular blessing to Academician Evgenii Velikhov, head of the Soviet nuclear fusion programme. Velikhov is a great advocate of international cooperation on fusion. It was he who put the proposal for the INTOR international tokamak project to the International Atomic Agency.

Vera Rich

US non-proliferation policy

Pact with China almost ratified

Washington

THE US Senate voted last week to approve President Reagan's controversial nuclear cooperation agreement with China, adding only some less-than-onerous certification requirements the President must meet to ensure that the agreement does not contribute to the spread of nuclear weapons. The Senate's approval is expected to be echoed shortly by the House of Representatives.

Opponents of the agreement have denounced it as a step backwards for the cause of nuclear non-proliferation: intelligence reports have been cited to support the claim that Chinese nuclear engineers have helped Pakistan to achieve a suspected nuclear weapons capability, and there have been reports of nuclear dealings with Argentina, South Africa and Iran. The agreement as signed by the President includes no formal non-proliferation safeguards, and critics say that this will encourage other countries to refuse to submit to non-proliferation safeguards.

The administration's answer is that China need not submit to safeguards because it is already a nuclear weapons state. The administration says that China has provided verbal assurances of its commitment to non-proliferation, but these have not been sufficient to quell anxieties in Congress. As recently as 1978, a Chinese leader publicly stated that all nations had the right to "nuclear strength".

More recently, China's position has changed: it has joined the International Atomic Energy Agency (IAEA) and in September announced for the first time that it would voluntarily open some of its civil nuclear facilities to IAEA inspectors. It will also apply IAEA non-proliferation provisions to its own exports.

Nevertheless, both the Office of Technology Assessment and the Nuclear Regulatory Commission have raised doubts about the agreement since it was introduced to Congress in July. Does it, for example, provide sufficient guarantees in the event of a serious deterioration of relations between the two countries? Although it specifies that prior mutual agreement would be needed before any US-derived nuclear fuel could be reprocessed or further enriched, many question how effective this would be if China wanted to process for weapons use stocks of nuclear fuel already on hand. And the agreement includes no clear US veto rights over exports of weapons-grade plutonium made from US fuel.

The sudden climb-down last week by opponents of the agreement seems to represent a simple recognition of political realities: to demand a substantially changed agreement that would require new negotiations with China would have needed an unlikely two-thirds majority in

Congress. Furthermore, any action would have had to be taken before 11 December, when the agreement is due to come into effect.

The modest requirements attached to Congress's approval are that the President must certify that nuclear technology and materials supplied to China will be used solely for their intended peaceful purpose, and that China will provide unspecified

"additional information" on its non-proliferation policy. In addition, the President must certify that US domestic law will be observed, and that requests for reprocessing approval can be granted or denied individually. The administration has cooperated with legislators on the wording of the certification requirements and is content with them; the way appears now to be open for US nuclear construction companies to compete with their European counterparts for contracts in the Chinese nuclear power programme.

Tim Beardsley

Television technology

Picture compression in sight

Princeton, New Jersey

SQUEEZING more television channels into a given bandwidth, every communication company's objective, is also high on the list of RCA's objectives. The company's research laboratories here are designing a novel electronic test unit that will double the number of television channels that can be carried by a satellite transponder.

The computer-controlled unit will show how well imperfect television signals, with some picture frames entirely missing, can be regenerated at the receiving end. To begin with, the tests will be conducted omitting every second frame of a complete television transmission. They will then be extended to see whether the system will work when more than half the signal is not transmitted.

The technique, called Motion Adaptive Interpolation, relies on a computer algorithm to reconstruct the missing frames. A less refined version of the system has been operating since 1983 on the telecommunications satellite, Satcom V, built by RCA to carry television signals to Alaska. Signals at the receiving end are regenerated from signals from which half the frames are missing, but the quality is poor. RCA considers that substantial improvements would be needed if the system were to be used for television traffic within the mainland United States.

RCA now operates five telecommunications satellites, and plans to launch two more by the end of the year. Roughly half the traffic consists of television transmissions between the major television networks and their affiliated stations and between programme suppliers and cable networks.

Predictably, processing television signals from which every other frame is missing imparts a jerky motion to moving objects. The machine on which RCA is working would store successive frames in such a transmitted signal and then, by interpolation between the positions of moving objects, reconstruct the missing intermediate frame.

If the development proves successful, RCA plan to use the system on its video-conferencing links and would also expect

suitable computing machinery to be installed at the head-ends of cable television networks, at the points where signals are received from linking satellites for onward transmission through the network.

Another spur to these developments is the general concern for the design of television receivers for the 1990s, on which several manufacturers are working. One likely development is that of digital transmission of the information required for each picture element (pixel), which among other advantages, will allow colour and luminance information to be transmitted separately, allowing greater control of picture quality.

Part of the development now under way is psychologically based; RCA wants to know how the human eye perceives motion in a sequence of two-dimensional images. The research group has provisionally worked out a scheme whereby it is not necessary to compare entire television frames to detect motion in one relative to the other. Instead, a sequence of progressively more rudimentary frames is constructed by omitting alternate rows and columns of pixels from a frame; the hope is that it will be possible to identify significant motion in the most rudimentary of these digital pictures, thus reducing the amount of data to be handled by the computers. It may then be feasible to incorporate enough processing capacity into individual television sets.

The same principle of data compression is being used in the development of a "smart" camera which may be used as a means of detecting only significant movement within its field of view.

Used as an automatic surveillance system, the camera is controlled by a modest microprocessor whose limited capacity is nevertheless able to detect movement in the most rudimentary of the sequence of collapsed frames and then to arrange to inspect a more informative frame in the hierarchy on the succeeding cycle of operation. RCA hopes that the US National Aeronautics and Space Administration will use the system as a motion detector in the gravity-free environment of the space cabin.

Bill Johnstone

Palaeoanthropology

All human life is there

EVEN the most hard-bitten hominid-fossil watchers were impressed by the specimens unveiled last week at the Commonwealth Institute in London. Centrepiece of the exhibition "The Human Story" is a sequence of candidates as human forefathers ranging from *Aegyptopithecus* and *Ramapithecus* through the australopithecines to *Homo* in his/her various forms.



An australopithecine, of the same type as Lucy, the skeleton found in Ethiopia by US anthropologist Don Johanson.

The skulls are boldly displayed so that the visitor can touch them, with mirrors beneath allowing examination of the teeth in a way that would make a dentist feel at home. Pride of place among the fossil exhibits goes to the 1.6 million-year-old *Homo erectus* skeleton WT15000, discovered last year by Kamoya Kimeu near Lake Turkana in Kenya and described in *Nature* in August (316, 788–792; 1985). Kimeu was on hand at the opening of the exhibition to explain that the skeleton on display is a replica — very convincing right down to the Kenya Museum accession numbers in black ink — and that the original is locked away in the Kenya National Museum, safe from the hazards of air travel and public exposure. Back in the museum too are more teeth, eventually destined to rejoin the rest of WT15000.

Another famous hominid fossil, "Lucy", is present in skeletal form and also as one of the life-size figures, reconstructed on the basis of the latest anatomical and cultural data. Lucy, or at least a

female *Australopithecus afarensis* from around 3.2 million years ago, is to be seen reaching up, hypothetically and bipedally, to gather fruit from a bush. It was a hard life for *A. afarensis*, it has been deduced from their fruit-eating ways and tendency to die off at an average age of 22, hence the wasted appearance of the reconstructed figure.

By the time the human line gets to 100,000 years ago, the reconstructed figure looks far meatier in the form of Neanderthal man. Not the clumsy colossus that has stubbornly survived in popular fiction for many years, but a character looking not unlike a modern-day middle-weight boxer, though one who has had a long career of mismatches.

The first question asked by many visitors to the exhibition seems to be "But

when did they become us?", and the exhibits and catalogue give a good guide to the various possible answers. But the question more likely to be in the visitor's mind at the end of the exhibition is "Has it finished?". After the impressive series of fossils and figures, and stone tools you can play with, comes a dash through recorded history and a bank of flickering television screens, with all four UK channels in quadruplicate or more, apparently representing the present. And then some conspicuously blank walls — the future is unknown, you see . . .

The exhibition owes much to sponsorship by IBM and in February it will take to the road, with stays in Amsterdam, Stockholm, Bremen, Paris and thence to Africa for display in Ghana, Nigeria, Senegal and Kenya.

Charles Wenz

"The Human Story" is at the Commonwealth Institute until 23 February 1985. Details from: Commonwealth Institute, Kensington High Street, London W8 6NQ, UK.

French education

Stoking the university fires

JEAN-PIERRE Chevènement, French minister of education, plans to double the number of potential university entrants in France by the year 2000.

This could make France an extraordinary place in which to be educated in the 21st century, for already there are a million students in French universities, compared with, for example, only 175,000 university students in Britain. In Britain, fewer than 5 per cent of the population are considered likely to benefit from university education; but in France, by the end of the century, 55 per cent of 17-year-olds could have the coveted "baccalaureat" or "bac" that would give them the right to enter university.

Whether they will enter university, however, is another question. Chevènement's plans include extending the bac — or bacs, as there are half-a-dozen different specializations — into more career-oriented and technical directions. Politically, he wishes to extend the opportunity of gaining the cachet of a "bac" to the largest possible number of children (at present only a third attempt it, a fraction he would like to increase to four-fifths by the year 2000); educationally, he wants to see a work-force better oriented to the demands of 21st century technology. Thus by 2000, the plan is that many who succeed at the bac will not be going to university, although they will have the right, but straight into work, in a new, high-technology France.

Hence if enough work exists in 2000 (at present many unemployed French school-leavers spend a year coasting in university), the universities need not be much larger by that date. But Chevènement plans to increase the number of qualified school-leavers to cope with the changing teaching

load, and a figure of 4,000 new teachers a year has been mooted. Some 400 new schools are also to be built.

Further, Chevènement is to loosen the stranglehold of mathematics on French education, much to the disgust of the mathematicians who consider that mathematical ability is essentially a test of intelligence. Thus the volume of mathematics teaching even in the coveted bac-C (mathematics and physics) is to decrease, and in other bacs it will become less abstract and oriented more to the needs of the central disciplines. (For example, in the "management" bac, bac-B1, mathematics will emphasize statistics and economic calculations.)

Among French scientists, reaction is mixed to the proposal to "dilute" mathematics. Biologists favour the move, claiming that many potential biologists are swamped in school by the mathematical demands of the science bacs; but physicists and mathematicians fear a fall in student quality. Chevènement, for his part, has no desire for the number of potential French Fields Medallists to fall.

Chevènement's planned school reforms cannot take effect before autumn 1986, however, by which time a right-wing government will probably be in power. Nevertheless, the minister's plans may not be forgotten. Many of his educational reforms, including bringing back the Marseillaise and improving discipline in primary schools, have gone down better on the right than the left. And Chevènement, once the darling of the left, has made no secret of the fact that he would accept an offer to continue his work, even under a right-wing prime minister. An imaginative conservative government might just agree.

Robert Walgate

US defence research

Will Congress cut SDI?

Washington

AN apparently unending series of exposures of fraud and waste in the Department of Defense (DoD)'s procurement division, coupled with anxiety over the federal budget deficit, is prompting Congress to take a tough line with defence spending this year. One of the casualties is likely to be the Strategic Defense Initiative (SDI), President Reagan's programme of research into anti-ballistic missile defences. Non-SDI research, however, is blooming.

The President's request of \$3,700 million for SDI in fiscal year 1986 (which began on 1 October) has been rudely refused by the House of Representatives, which has offered a mere \$2,500 million. The Senate has not finally made up its mind, but the figure of \$2,960 million has been urged by a committee. (The figure in 1985 was \$1,400 million.) Despite the apparently large increase from 1985, the SDI organization is far from happy, and has made plain to Congress that support at the level that now seems likely will lead to delays of a year or more to key demonstrations.

But if strategic defence research is now in question, other basic research supported by DoD is likely to increase substantially in 1986. This budget item request (known as line 6.1 to insiders) was \$971 million, which is likely to be met or even exceeded. Last year the figure was \$860 million. Historically, roughly half of the 6.1 budget, and some of the more applied 6.2 research, has been spent in universities. The increase for basic research continues a trend established more than ten years ago. This year, however, the Department of Defense has a new pot of money for university research that it is asking to have filled, the University Research Initiative (URI). Responding to widespread concern about the state of science education and research at US universities, the department went to Congress to ask for \$25 million for URI in 1986; the programme would support fellowships and the like at DoD laboratories as well as university research in risky but potentially profitable areas such as materials and structures, fluid mechanics, biotechnology, communications and optical networks. Congress was so impressed with the proposal that a Senate committee voted a total of \$100 million for URI (and DoD has quickly decided that it will, after all, be able to use the lot). This is on top of a \$30 million per year university instrumentation programme and \$500 million of direct research support last year. Even Colonel Donald Carter, acting head of DoD's research and advanced technology division, one of whose jobs is to drum up support for DoD university research on Capitol Hill, admits that

"we're doing quite well this year". But the House of Representatives has so far agreed only to the original request for \$25 million; a compromise will be worked out with the Senate next month.

Universities, despite their perpetually uneasy relationship with DoD because of friction over the issue of research secrecy, are not unnaturally delighted that DoD is "taking the university research infrastructure seriously", according to Robert Rozenzweig, president of the Association of American Universities. But the high-principled association still objects to the addition of inoffensive sounding amendments to other legislation, including DoD appropriations, that seek to provide funds for specific research projects or facilities at particular institutions, usually within the constituencies of the sponsors of the legislation.

Pork-barrel legislation has a distinguished history in the United States. Some 34 projects were supported in this way between 1983 and 1985, according to data

compiled by the association. But Rozenzweig and Robert Clodius, president of the National Association of State Universities and Land-grant Colleges, say that if allowed to proceed unchecked, direct funding of research facilities which circumvents peer-review procedures will "undermine the very system that has made our research enterprise the envy of all other nations". The presidents are lobbying congressional committees to eliminate all pork-barrel amendments.

Among the beneficiaries of current proposed pork-barrel amendments (in a number of different bills) are: Syracuse University (\$12 million); Oklahoma State University (\$1 million); Rochester Institute of Technology (\$11.1 million); Northeastern University (\$13 million); University of Nevada (\$3.5 million); University of South Carolina (\$4 million); East Michigan University (\$2 million); University of Missouri (\$450,000); New York University (\$2.6 million); Tufts University (\$1 million); and Pennsylvania State University, the University of Minnesota and Massachusetts Institute of Technology (each to receive one-third of \$1.7 million).

Tim Beardsley

Japanese audiotechnology

Renoir returns from the dead

Tokyo

If you want to hear about the Impressionist school of painting, who could be better to listen to than Pierre Auguste Renoir who, along with Monet, was its most famous exponent? The only problem is, of course, that Renoir died in 1919. But thanks to a little modern computer technology this proves to be a trifling objection — at least in Japan.

To help advertise a major exhibition of the Impressionists being held in Tokyo until the middle of December, callers can hear about the philosophy of Impressionism from Renoir himself by dialling Tokyo 320-3000. The voice is the product of the Japan Acoustics Research Laboratory and its construction relies both on the ability to analyse and manipulate sound patterns with the aid of computers and the ability to predict fundamental voice characteristics through knowledge of an individual's anatomy.

According to the institute's director, Dr Masumi Suzuki, many of the major characteristics of a voice are governed by the structure and shape of the oral and nasal cavities and the resonances they produce in air set vibrating by the vocal cords. Of course, regional accents and the like are not shaped only by such parameters, but the characteristics that enable one to recognize a voice independently of its accent are. The structure and shape of the oral and nasal passages can themselves be measured relatively easily from an X-ray. Where such data are not available, they can be predicted with less accuracy from de-

tailed measurements of the face and neck.

Once the details of the vocal tract are available, simulation proceeds through use of a computer model. A human voice is input (reading French in Renoir's case) and analysed spectrographically. A computer model of the effect of the different resonances produced in different regions of an individual's vocal tract is then used progressively to modify spectral components of the input voice. The result, in Renoir's case, is a new voice that still, according to native French speakers, is perfectly accented French but has (lacking evidence to the contrary) acquired the characteristics of Renoir's voice.

But the "science" of voice analysis does not stop there. If a voice can be predicted from a face, why not a face from a voice? Some researchers have had a try with one of Japan's most wanted criminals, the self-styled "man with 21 faces" who extorts money from confectionery manufacturers by placing poison in bars of their chocolate on sale in supermarkets. All that is known about him is the sound of his voice on tapes and recorded calls. Despite the production of a portrait based on his voice, however, his arrest has moved no closer.

Patriotic Americans are not forgotten either in famous historical figures whose voices have been brought back from the dead. One can hear Abraham Lincoln proclaiming "Fourscore and seven years ago our fathers brought forth on this continent a new nation . . .", in tones to which his high and slightly bent nose apparently contributed a great deal. **Alun Anderson**

UK computing

Prospects good and bad

By the United Kingdom's own rather sad standards, computing research in universities is doing relatively well. Sir Keith Joseph, Secretary of State for Education and Science, this month gave the Computer Board for Universities and Research Councils an extra £1 million for 1986-87, to be followed by £2 million for each of the next two years, to enhance Janet, the wide area network that will connect computers in all universities, research council sites and polytechnics. Although the demand by researchers for computing facilities far exceeds resources, the Computer Board continues to make a relatively successful case to the government for increased support, most recently in its three-yearly report (see below).

In the private sector, computer companies themselves are finding life difficult, but the manufacturing industries are greatly increasing their use of computers. A survey of more than 2,000 engineering companies, sponsored by the Department of Trade and Industry and published this week*, shows that nearly 60 per cent of

UK engineering plants are using or are about to use computers. In 1986, the industry plans to spend £250 million on software and £600 million on hardware, mainly on minicomputers and mainframe computers rather than micros. John Butcher, Under Secretary of State for Industry, says that the United Kingdom may be leading Europe in the uptake of computers for engineering purposes, "although European data are hard to come by".

The most widely used applications of computers in industry are in manufacturing management and mechanical design, although computer-aided design and draughting are the fastest-growing applications. The top-selling microcomputers are IBM personal computers, XTs and ATs; Apricot is a poor second with only half as many installations as IBM.

IBM is the largest manufacturer of computers in the United Kingdom, and, according to a forecast published last week*, although the company's activities contribute to import savings, UK computer companies are growing too slowly. "The prospects for ICL [Britain's nationalized computing company] are less clear now that the government has largely abandoned its 'buy British' policy and the outlook for small computer companies is extremely volatile", according to Cambridge Econometrics' model. The deficit on the balance of payments for office machinery and electronic data-processing equipment rose from £200 million to £1,100 million between 1980 and 1983 (current prices).

Cambridge Econometrics expects industry's investment in computing equipment and export demand to increase by 4 per cent per year for the next 10 years; by the end of the century, the demand for office machines and data-processing equipment will still be the highest for all manufacturing industry in Britain. But British companies are unlikely to be taking advantage of these opportunities.

The study suggests that new domestic and foreign entries to UK production during the next few years may improve the viability of the industry, but it says that "we do not see enough evidence at the moment to build such hopes into our central forecast". Large companies at present involved in other areas of electronics, particularly military applications, could enter the civilian computer market, or more foreign companies could locate European production centres in Britain. But these remain hopes rather than predictions.

Maxine Clarke

**Third Annual Engineering Computers' Survey in Engineering Computers*, November 1985; *Prospects for Computers and Office Machinery in the UK to the year 2000* Cambridge Econometrics forecast 85/2. Price £500.00 from PO Box 45, 21 St Andrews St, Cambridge CB1 2EX, UK.

UK board reports

The report of the UK Computer Board for Universities and Research Councils, published this month (Cmd 9671, HMSO, £3.30), covers the period April 1983 to March 1985. During this time, the board was mainly concerned with supporting local computing facilities for research and teaching in universities, supporting national facilities and promoting the efficient sharing of resources. The most important developments were the extension of supercomputing facilities in the two national centres and the development of Janet, the wide area network connecting all research establishments. The board has also taken a lead in developing standards and encouraging the use of computers for teaching in universities. Reappraisal of some of its methods has been caused by its observations of developments in the United States and Japan.

During the next few years, the board will re-examine its relationship with the University Grants Committee (UGC) and will assess its priorities for what it sees as the two ends of a spectrum — supercomputers for the specialized research user and workstations for the undergraduate user.

A joint working party with UGC and the Advisory Board for the Research Councils has set up two subgroups; one, chaired by Professor A. MacFarlane (Cambridge) will examine the scientific need for advanced computing facilities and the other, chaired by Professor E. Spratt (Kent) will examine technical options for meeting that need, particularly in collaboration with industry.

Maxine Clarke

UK space centre

Space programme coordinated

BRITAIN last week re-entered the international space arena with the announcement of its National Space Centre (NSC), to be based in London. The first director general of the centre is Roy Gibson, who was also the first director of the European Space Agency (ESA), for a three-year period. NSC's initial budget is £100 million a year, or about one-fifth of the French government's space budget.

The purpose of the centre, a concept that the Minister for Industry Geoffrey Pattie has taken the lead in promoting, is to coordinate Britain's fragmented space programme, which at present includes projects sponsored by the Department of Trade and Industry (DTI), Ministry of Defence (MoD), Science and Engineering Research Council (SERC) and Natural Environment Research Council (NERC). But "no change in policy" has occurred with respect to ESA; NSC is intended to make Britain's contribution "more efficient".

The first task for NSC is to develop a long-term strategy for Britain's space effort by coordination of industrial, defence and scientific users of space, who will be able to influence the policy of the centre by "financial and in-kind contributions". NSC's first projects will be based at the Royal Aircraft Establishment, Farnborough (where DTI and MoD already have programmes) and at SERC's Rutherford Appleton Laboratory, the present focus of the academic space programme in Britain.

Professor Bill Mitchell, chairman of SERC, welcomed the formation of NSC last week on behalf of SERC and NERC. The main interests of the research councils are astronomy from outside the atmosphere and remote sensing. SERC, NERC and DTI already collaborate on the latter.

The centre will not be involved in the US Strategic Defensive Initiative "at this stage" — the largest proposals with which it is at present concerned are whether, as a member of ESA, to take part in the US space station planned for the 1990s and whether to support the development of the novel reusable launch vehicle *Hotol* in collaboration with the private sector.

Britain pulled out of Europa, its collaborative space venture with France and Germany, 25 years ago. Since the failure of that project, France has taken the lead in the development of Ariane, now a commercial competitor of the US space shuttle. Although Britain's effort has a long way to go before catching up with other countries with space programmes — not only France and the United States, but also Japan, West Germany, Italy and India — the creation of NSC is a step in the right direction.

Maxine Clarke

AIDS in Japan

No screening of blood donors

Tokyo

OVER the centuries, Japan has fought off foreign invaders with a fervour second to none, but in the face of an invasion of AIDS (acquired immune deficiency syndrome), Japan has been slow to protect the purity of the national blood. The AIDS virus has already made significant inroads into the blood of certain sectors of the Japanese population. Yet there are still no plans to introduce routine screening of blood donors as in the United States and Britain.

The number of confirmed cases of AIDS in Japan has risen to eleven since March, when the first two cases were reported among haemophiliacs (*Nature* 315, 8; 1985). Six of the eleven were or are homosexuals and five haemophiliacs. Two of the homosexuals and four of the haemophiliacs had died by late October according to the Health and Welfare Ministry.

Many more Japanese are known to have been exposed to the virus. In September, a team at Jutendo University headed by Takao Matsumoto found that the blood of 5 out of 103 homosexuals contained antibodies to the AIDS virus, HTLV-III (human T-cell lymphotropic virus), while in October the Health and Welfare Ministry revealed that, of 395 haemophiliacs surveyed across the nation between September 1984 and this summer, 120 or 31 per cent have been exposed to AIDS, probably through treatment with blood plasma products from the United States. With about 5,000 haemophiliacs and 300,000 homosexuals in Japan, estimates of those already exposed to AIDS run into tens of thousands.

In response, the Health and Welfare Ministry has announced plans to double the amount of blood taken from individual donors to reduce dependence on imported blood plasma (*Nature* 317, 101; 1985), but earlier this month, in reply to an inquiry by a communist member of the House of Representatives, the ministry said there is at present no need for mandatory AIDS screening of blood donors. Instead, "voluntary" testing will be covered by the health insurance system for "high-risk" groups such as homosexuals and drug addicts. But in Japan, where pressure to conform is severe, a rush of volunteers can hardly be expected.

Blood donors in Japan are routinely tested for hepatitis virus and venereal disease, which may have induced homosexuals to give blood in order to get free VD checks. A month after the release of the findings by the team at Jutendo University, the Health and Welfare Ministry directed the Red Cross Society to stop accepting blood from homosexuals.

Amid calls for mandatory AIDS tests for blood donors come appeals for the

introduction of screening for another less notorious but equally insidious virus, HTLV-I, which causes ATL (adult T-cell leukaemia). Yorio Hinuma of Kyoto University's Institute for Virus Research estimates that a million Japanese are now carriers of ATL virus. Although the normal infection route is from husband to wife and from mother to child, Hinuma suspects that each year as many as 40,000 Japanese may be "artificially" infected with ATL virus by blood transfusion. Only one in a thousand carriers usually develops the disease but for those who do, ATL is almost invariably fatal, with 50 per cent succumbing within six months and the rest within two years. Professor Hinuma stresses, however, that there is very little relation between the ATL virus HTLV-I and the AIDS virus HTLV-III, and he considered the HTLV-I, II, III terminology proposed by Dr R.C. Gallo, chief of the tumour cell laboratory of the US National Cancer Institute, to be inappropriate.

As in the case of AIDS, the Health and Welfare Ministry has turned down requests for mandatory screening of blood donors. In both cases, the cost of screening the 8 million people who donate blood in Japan each year appears to be the major deterrent.

David Swinbanks

More AIDS money

Washington

ALARM over the spread of acquired immune deficiency syndrome (AIDS) was evident in an agreement of funding for the US Public Health Service reached last week in Congress. The report established a special fund of \$70 million for combating the disease to be allocated by the director of the National Institutes of Health (NIH); this is in addition to the \$120 million the administration had requested for AIDS research at NIH. The total sum appropriated under the agreement would be \$5,501 million, an increase of 7 per cent over last year's figure, with 6,100 new and competing extramural research grants — substantially more than the 5,000 the administration wanted. The total support for AIDS research throughout the Public Health Service is expected to be more than \$300 million.

The agreement, which has not yet been formally accepted by either arm of Congress, notes that there is an urgent need for more basic research on AIDS and directs NIH to spend "a substantial portion" of the extra \$70 million on basic research. The director of NIH is urged to convene a special working group of both intramural and extramural researchers to decide priorities for such research, which would also report to Congress.

Tim Beardsley

Soviet spaceflight

Medical hazards to cosmonauts

THE illness of Vladimir Vasyutin, which led to the premature return to Earth of the latest long-stay crew of Salyut-7, has been provisionally diagnosed as "an inflammatory disease that did not respond to the drugs available on board". Although Vasyutin was unable to act as commander during the landing, he was well enough to give a brief television interview after touchdown. The doctor who gave him a preliminary check-up said that his condition was "satisfactory", but that he would nevertheless need hospital treatment.

Valerii Kybasov, himself a cosmonaut, told a TASS correspondent that it was decided to bring Vasyutin back to Earth simply because "in our country, man comes first". There have been some hints in the Soviet media, however, of fears that the malady might have spread to his fellow crew members, Viktor Savinykh and Alexandr Volkov. An unnamed official, interviewed at the landing site by Moscow radio, spoke with some concern of an ordinary cold having developed "into some new forms", requiring "urgent intervention from Earth. Both Savinykh and Volkov proved to be in good health, however.

This is the first major medical incident in almost 25 years of Soviet manned spaceflight, although there have been cases of toothache and colic which responded to the on-board medicine chest. Even if Vasyutin's indisposition proves relatively minor, his hospitalization could have important repercussions for the Soviet space programme. The lobby within the Soviet space programme pressing for a flight to Mars claims that the long-term Salyut missions have already established the human organism's adaptation to prolonged weightlessness without permanent after-effect. Vasyutin's illness will provide the "conservatives" who oppose the lobby with a powerful example of their main counter-argument, that space medicine has not yet reached the stage where cosmonauts can be allowed to venture out of reach of a base hospital.

Vera Rich

● Two Syrian trainee cosmonauts have arrived in the Soviet Union to prepare for a future joint mission. This news has caused some speculation among Baikonur-watchers, since even a few months ago, Soviet space officials were emphatic that no further "international" flights were being planned. Even more curiously, the news was announced in *Krasnaya Zvezda*, the Red Army daily. It has been suggested, therefore, that the invitation to Syria to participate in such a flight may be a sop, not unrelated to the tentative political rapprochement of the Soviet Union towards Israel. □

British environment research

From penury to mere poverty?

THE Natural Environment Research Council (NERC), the most friendless of Britain's four science research councils, is hoping that the tide of misfortune is beginning to turn. Mr Hugh Fish, the council's chairman, said earlier this week that he hopes, in the next two weeks, to have assembled the funds (\$2.5 million a year) that will enable Britain to join the new international Ocean Drilling Program being organized by the US National Science Foundation (NSF). The council has also won a modest success in the commercial exploitation of the sea-bed profiling system called GLORIA and hopes that a modest share of the extra £14 million made available for the research councils two weeks ago will allow it to increase its support for university research.

British membership of the Ocean Drilling Program (ODP) has been a thorny issue for most of this year. The hope, now, is that the necessary sum will be raised by means of contributions from industry (mainly the oil companies), government departments and from NERC's own budget. According to Mr Fish, a British delegation has been invited to the next meeting of ODP, in Hawaii in the second week of January, as observers, but he hopes they will be able to attend as paid-up members.

NERC's success with GLORIA (which stands for Geological Long Range Inclined Asdic) rides on a contract with the US Geological Survey last year for the survey of the bottom topography of part of the US economic offshore zone. Now, according to NERC, the contract has been extended to include virtually all of the offshore United States under a six-year contract that will cover, among other things, the development of improved equipment at the Institute of Oceanographic Sciences. At the same time, NERC has arranged to "capitalize on the opportunities presented by this unique British research module" by means of a contract with Marconi Underwater Systems, which have the right to build versions of the asdic system and to provide an underwater survey system overseas. The company apparently expects there to be some £5 million worth of business each year, and is committed to share some of the benefits with NERC.

A statement put out on Monday with the council's annual report for 1984-85 says that the decline of the value of research commissioned by government departments from council institutes, the chief cause of NERC's money problems in recent years, has now been reversed; last year, income of this kind amounted to £24.9 million, compared with £23.7 million in the previous year. Mr Fish said that he hoped the figure will increase as further opportunities for commercialization arise.

But NERC's hopes of winning public support for the costs of restructuring, the research council euphemism for persuading people to retire early, were disappointed; it asked for £2 million, but received only £1.25 million.

Mr Fish and his senior colleagues on Monday repeatedly emphasized how important it is that NERC's research should be commercialized. By means of a sales office in Brussels and an agent in Washington, the council hopes to win extra business from overseas companies and governments, "which is not to suggest that British government departments will not always be important customers". But there are obvious benefits in diversity. "We've had our fingers burned", said one official.

Meanwhile, NERC plans to introduce a more formal system for the assessment of its own activities by the three criteria of timeliness; potential economic and social benefit to the United Kingdom; and the enhancement of Britain's reputation overseas. There is to be a process by means of which the outcome of research decisions will be judged retrospectively by these criteria, with the objectives both of evaluating decisions made and of refining the criteria.

For the immediate future, NERC hopes to increase its support for university research, by about £5 million a year on its present spending, estimated at close on £17 million, counting the cost of research vessels maintained by NERC. Mr Fish says that university research is especially desirable because of its flexibility; people will do "basic research on a short-term basis".

Dr John Bowman, secretary of the council, also said on Monday that there was a case for increasing NERC's interests in atmospheric chemistry, partly so as to make good the loss of support for acid precipitation research some years ago but also for more general reasons. The council also plans "a more sophisticated approach to forestry" and says it has already written to the new director of the National Space Centre (see p.305) to declare its interest in remote sensing of various kinds.

On the research council's future organization, Mr Fish said that a new version of its corporate plan was about to be shown to members of staff for their comments, and would be finally discussed by the council in January. On this occasion, he said, the plan would be more concerned with science than with organization, which would nevertheless not be substantially changed from that in the document published earlier this year.

The council is pushing ahead with the appointment of three scientific directors to oversee the different parts of the council's programme; one is in post and two are

about to be appointed. But it may be decided that these persons should be sited in laboratories and not in the council's headquarters. The future of the British Geological Survey, the largest of the council's dependants, would be decided only after the committee on the subject under Sir Clifford Butler had reported, although Mr Fish gave it as his opinion that the survey needs more money to support its "core" programme.

Last year, the council's total income was £93.4 million, of which £65.3 million was contributed by the general science budget, which supports all the research councils. Although contract research is seen as a growing component of the budget, Mr Fish says that NERC has now learned a lesson from the painful withdrawal of government contracts in recent years, and intends not to employ full-time members of its staff on short-term contract work. □

French experiment on coma patient

FRENCH doctors experimented on a patient who had been in a coma for three years, it was reported in France last week amid a storm of protest. The patient, a young man who was claimed to be "brain dead", has since died, but not, it is claimed, as a result of the experiment, which involved a high-speed blood transfusion into a pelvic bone.

The doctors are reported to have undertaken and announced the experiment to provoke debate on the ethics of such operations, and this objective they have certainly achieved. Thus Louis René, president of the ethical commission of the French medical association, has said he was "profoundly shocked" to hear of the experiment. "Since 1947 at Nuremberg — a symbolic date and place — doctors have affirmed that experiments can only take place with the agreement of the subject", René wrote to the newspaper *Le Monde*. In the case of a patient in coma, this agreement is patently impossible and so experiments on comatose patients should be forbidden, he wrote. Such experiments "treated the patient as an object and denied him all human character", exemplifying "condemnable medical imperialism".

Following this line, Edmund Hervé, minister of health, has demanded an inquiry and cited article 18 of the French medical code which forbids a doctor to subject a patient to unjustifiable risk.

Le Monde, however, quotes a somewhat different view from M. Guy Tiercelin, president of a national anaesthetics commission. "I am not necessarily opposed to an experiment if there is proof of brain death", said Tiercelin. "One could test certain techniques of reanimation, but one must be very vigilant, for we should never perform experiments on patients who may recover".

Robert Walgate

Japanese psychiatry

SIR—The article by Alun Anderson entitled "Abuse for visiting scientists" (*Nature* 315, 361; 1985) contains some noteworthy misunderstandings, and as one of the parties concerned we would like to clarify matters.

Protest action taken against the seventh congress of the Japanese Society on Biological Psychiatry at Gifu was not directed at the visiting scientists from the United Kingdom and the United States. The protest was made in order to obtain open public discussion on the ethics of an experiment on a human fetus carried out by Professor Masayuki Namba of Gifu University and his followers. Professor Namba, the president of the congress, was responsible for the human experiment in his capacity as director of the Department of Neurology and Psychiatry in Gifu University. But he chose to avoid discussion and to cancel the congress.

The Japanese Society on Biological Psychiatry is managed under a highly closed, almost secret system. For example, a non-member who wished to join only for the days of the congress at Gifu had to obtain a written recommendation from one of the board of directors, even though the names of directors and the address of the office are not usually announced. The society has no ethical code governing experiments involving human subjects. We have examined the summaries of papers that were to be delivered at the Gifu congress and past congresses and find evidence of many unethical experiments. So we would like to give a warning about the true nature of the society. The issue of the human experiment at Gifu is now under investigation by an *ad hoc* committee of the Japanese Society of Psychiatry and Neurology but we understand that the doctors involved in the experiment are unwilling to cooperate.

We, the Psychiatrists' Union of Tokyo University, aim to reform the present state of psychiatry in Japan and to that end will take action against unethical psychiatric conduct. A liaison group, including some of our members, has been pressing for open discussion about the Gifu experiment. This group is not, as Anderson wrote, the "antipsychiatrists within Tokyo University". Nor did this group intend to visit the smaller meeting at Nagoya. It is also not correct to describe us as "antipsychiatrists" as there are differences between our views and those of the antipsychiatrists found in the United Kingdom and Europe in the 1960s and 1970s.

Anderson states that there is no evidence of unethical conduct by the outpatients ward doctors of the University of Tokyo hospital. Why is he so dogmatic? Anderson has mentioned the horrific incidents that occurred at Utsunomiya hospital. But the outpatient doctors have long been closely connected with the Utsu-

nomiya hospital as has been widely reported in the Japanese press.

For example, the director of the Utsunomiya hospital set up an outpatient clinic near Tokyo University hospital and the outpatient doctors made use of this clinic for practice and research. Some patients were sent from this clinic to the Utsunomiya hospital. It is a fact that many of the outpatient doctors visited the Utsunomiya hospital to work or for research and contributed to the hospital's development. Many further connections exist.

The miserable state of Japan's mental hospitals as mentioned in Anderson's article is a social problem for the nation. It stems partly from the relationship between university mental hospitals as suppliers of new doctors and other mental hospitals. Many university hospitals treat the mentally ill as research objects and think little of their rights. This trend is closely related to the abuses seen at Utsunomiya where patients were forced to remain in hospital for long periods and were dominated by violent methods.

Furthermore, the government supports this situation through the Japanese Mental Health Act because few legal rights are given to patients, especially those admitted to mental hospitals. Visits from Disabled Peoples International, the International Commission of Jurists, and the International Commission of Health Professionals are necessary and important. These visits are expected to have a favourable influence upon mental health policy in Japan. We regret to say, however, that a large number of Japanese psychiatrists take attitudes similar to the outpatient doctors of Tokyo University. Reform is still far off. To be reconciled with the outpatient doctors is not easy for us at the present.

KIMIO MORIYAMA
(Acting Chairman)

Psychiatrists' Union of Tokyo University,
7-3-1 Hongo, Bunkyo-ku,
Tokyo, Japan

Slighted Dallas

SIR—In his appreciation of the 1985 Nobel Prize award in Physiology or Medicine to Michael S. Brown and Joseph L. Goldstein (*Nature* 17 October, p.569), Peter Newmark has indulged either in the subtlest of *Nature's* ironies or an insult that cannot have been intended by so sympathetic a commentator. Newmark points out that "So far all attempts to lure both of them from their opulent Dallas surroundings to more distinguished institutions have failed". As a longtime admirer not only of our new laureates but of the remarkable medical school in which they work, permit me to suggest that it might be hard, indeed, to identify "more distinguished institutions". Considering the high level of innovative science that is con-

ducted at the University of Texas Health Services Centre in Dallas, I look forward to Newmark's revised estimate of Dallas as more of her sons share the dais with the King of Sweden.

GERALD WEISSMANN

New York University Medical Center,
550 First Avenue,
New York, New York 10016, USA

Irrigated tomatoes

SIR—I would advise Michael Andrews (*Nature* 17 October, p.570) not to drink water containing 3,800 micrograms of selenium per litre. As Paracelsus said, "the dose alone determines the poison", and a woman who took about 27,300 micrograms of selenium daily for 2 months had toxic symptoms¹. No increase in symptom and disease reporting was found in a population drinking water containing 494 $\mu\text{g l}^{-1}$ of selenium². Andrews does not report the selenium content of the tomatoes he filmed in July, which he wanted to eat. If they were irrigated with water from the California Aqueduct, the water was reported³ in the same month to contain 0 to 2 $\mu\text{g l}^{-1}$. Information on selenium content of common foods is supplied by Underwood⁴.

THOMAS H. JUKES

University of California, Berkeley,
Department of Biophysics,
Berkeley, California 94720, USA

1. US Public Health Service, Morbidity and Mortality Reports, 33, 157-158 (1984).
2. Valentine, J.L., Kang, H.K. & Reishord, L. Abstracts, 3rd Int. Symp. on Selenium in Biology and Medicine, Beijing, 26 May 1984, p. 139.
3. Kennedy, D.N. Report of Activities of the Department of Water Resources, Sacramento, California 20 July 1985.
4. Underwood, E.J. (ed.) *Trace Elements in Nutrition* (Academic, New York, 1977).

Evolving journal

SIR—Your review of *Molecular Biology and Evolution* (*Nature* 26 September, p.296) contains two errors. In one, your reviewer asserts that *Molecular Biology and Evolution* "has not expanded since its first issues". The five issues of 1984, the special inaugural issue of December 1983 not being included, totalled 366 pages or 73 pages per issue. The six issues of 1985 totalled 560 pages or 92 pages per issue. That is a 25 per cent increase. That, together with the fact that, in its first year, *Molecular Biology and Evolution* circulation surpassed that of the excellent 14-year-old *Journal of Molecular Evolution*, suggests, contrary to your reviewer's report, that *Molecular Biology and Evolution* has already established itself alongside its worthy competitor.

WALTER M. FITCH

(Editor-in-Chief,

Molecular Biology and Evolution)

Department of Physiological Chemistry,
528A Service Memorial Institute,
University of Wisconsin,
1300 University Avenue,
Madison, Wisconsin 53706, USA

Testing for cystic fibrosis

A start has at last been made on the development of a prenatal test for cystic fibrosis. At first, it will only be useful in some families.

A BATCH of papers published in this issue of *Nature*, together with a paper in the current issue of *Science*, signal the beginning of the end of the hunt for the cystic fibrosis gene, defects of which are responsible for the disease. Although the gene remains to be identified, at least it is now certain that it resides in the middle of the long arm of chromosome 7, within a region that comprises no more than about a million base pairs. And the new data offer immediate prospects for the prenatal testing for cystic fibrosis.

Approximately one in twenty Caucasians carry a defective cystic fibrosis gene, and about one child in 2,000 inherits such a gene from both parents and consequently dies of the disease, usually by the age of 30. Despite the frequency of the disease, nothing has been learnt of the gene (or genes) in which the defects arise and no gene product has been convincingly demonstrated. It has not therefore been possible as yet to devise a prenatal screening test for cystic fibrosis relying on the detection of the defective gene or its product. An alternative approach, laid out in detail by Botstein *et al.* (*Am. J. hum. Genet.* 32, 314; 1980), is to look for and find a restriction-fragment linked polymorphism (RFLP) close enough to the cystic fibrosis gene to act as a linkage marker.

An RFLP is no more than a variation in DNA sequence that generates or abolishes a recognition site for a restriction enzyme. Because these enzymes cut through double-stranded DNA, the positions of their recognition sites on a stretch of DNA determine the sizes of the "restriction fragments" of DNA they generate. The chromosomes of any homologous pair are bound to differ by many RFLPs, which can therefore, in principle, be used to tell which of the pair of chromosomes carries, for example, a defective cystic fibrosis gene. If the RFLP is located closely enough on the chromosome to the cystic fibrosis gene not to become separated from it by the normal process of genetic recombination, it will also serve as a linkage marker for the defective gene. Because the RFLP is only a linkage marker, rather than a direct assay of the defective gene, it cannot be used for population screening but only predictively, in family studies.

It has been calculated that if RFLPs are more or less evenly distributed along and among the chromosomes and are

"informative", no more than 150 would be sufficient to provide one closely linked with any gene. So far, however, although several hundred useful RFLPs have been found and many have been localized, they are not evenly distributed and most are informative for only a minority of families. In searching for a linkage gene that could be on any chromosome, it therefore makes sense to begin with the chromosomes that are most saturated with RFLPs and conventional markers (such as different forms of serum enzymes). In this way, by the middle of 1985, several research groups had excluded about 40 per cent of the human genome as the site of the cystic fibrosis gene.

The first sign of success came at the Eighth International Gene Mapping Workshop in Helsinki in August where Hans Eiberg, of the University Institute of Medical Genetics in Copenhagen, reported a tentative link between cystic fibrosis and a form of the serum enzyme paraoxonase (an aryl esterase). Eiberg calculated that the paraoxonase gene and the cystic fibrosis gene lie within ten million base pairs of each other. However, this linkage — now published by Eiberg and others in *Clinical Genetics* 28, 265; 1985 — was less useful than it might have been because paraoxonase is poorly characterized and the chromosomal localization of its gene is unknown. Rather than dampening the enthusiasm of those who had failed to find an RFLP linkage, Eiberg's result revived their flagging energy.

Among those searching for an RFLP were Lap-Chee Tsui and colleagues at the Hospital for Sick Children in Toronto. In July of this year they came to an agreement to test some of the 200 or so useful RFLPs generated by Collaborative Research Inc. Much to their joy, and purely by chance, a cystic fibrosis-linked RFLP turned up among the first half a dozen RFLPs that Tsui tested. The result was reported at the American Society of Human Genetics meeting in October, and has just been published in *Science* (230, 1054; 1985).

What was not reported at the meeting was the chromosomal localization of the RFLP, information that would immediately have enabled others to concentrate exclusively on any RFLPs that they knew to be associated with the chromosome in question. Members of both the Toronto and Collaborative

Research teams say that they had only preliminary evidence at the time and were not prepared to risk presenting a faulty localization in public. Nevertheless, rumour emerged that their RFLP was on chromosome 7, as they confirm on page 380, although other rumours pointed towards different chromosomes.

Spurred on by the rumours, two other research groups have now found RFLPs that are linked to cystic fibrosis and report them on pages 322 and 384. One of the groups is centred on Bob Williamson's laboratory at St Mary's Hospital Medical School in London. Its RFLP probe is an "anonymous" DNA probe on the long arm of chromosome 7. Williamson's group also reports linkage between the cystic fibrosis gene and two other chromosome 7 genes — that for the T-cell receptor beta chain (p.384) and a collagen gene (*Lancet* 29 November p.1241). The linkage reported by the other group, centred on Ray White's laboratory at the Howard Hughes Medical Institute in Salt Lake City, is to an RFLP that is in the locus of the *met* gene, which is a human oncogene on the long arm of chromosome 7 (see p.285). The overlap between the various probes locates the cystic fibrosis gene in the middle of the long arm of chromosome 7.

Both Williamson's and White's RFLPs are much more tightly linked with the cystic fibrosis gene than is Tsui's. His RFLP is probably no closer than 15 million base pairs (0.5 per cent of the human genome) to the gene. Williamson's and White's are each within about a million base pairs of it, conceivably very much closer. The greater the distance between the RFLP and the gene, the greater the chance that they will be separated by recombination during meiosis, confounding attempts to use the RFLP as a prenatal diagnostic marker of the defective gene. In these terms, Tsui's RFLP is not diagnostically useful, whereas the other two are. What is likely to emerge, when the dust settles, is a set of markers which will make the reliable prenatal diagnosis of cystic fibrosis a reality assuming that all cases are due to a mutation of the same gene.

Both to develop a screening test that avoids the need of family studies and to discover the fundamental basis of cystic fibrosis, vigorous attempts will now be made to identify the gene itself and to find out what it produces. **Peter Newmark**

Evolutionary theory

Hamilton's rule OK

From Alan Grafen

THE darwinian principal of natural selection was extended by Hamilton in 1964 to the evolution of social behaviour¹. This extension is central to much of today's evolutionary biology, but it is something of a sport among mathematical biologists to cast doubt on its validity. Now David Queller, writing on page 366 of this issue², presents a simple and powerful derivation of Hamilton's result that should go some way towards convincing the sceptics.

Hamilton's extension to darwinism is enchantingly simple and has come to be known as 'Hamilton's rule'. In an uncomplicated world, the morphology and behaviour of an individual would affect the number of offspring it had in its lifetime but not the number of offspring other individuals had. Given that individuals do interact, though, the question is how will selection cause animals to value each other's reproduction? Will they place no value on it — that is, will selection favour forms according only to their own number of offspring? Or will they place equal value on their own reproduction and that of others — that is, will selection favour those forms that produce more offspring in total for the species? Hamilton's answer was intermediate between these two ex-

tremes. He showed mathematically that natural selection would act so that one individual would value the reproduction of another according to how closely the two individuals were related. Identical twins would value each other's offspring equally with their own; brothers and sisters would value each other's offspring half as much as offspring of their own; and so on. The evolution of a social trait then depends on an 'inclusive fitness' calculation, in which all the effects of the trait on reproduction are added together, weighted by the appropriate relatednesses. Relatedness matters because relatives share more genes than non-relatives, and natural selection is a matter of the nonrandom survival and proliferation of some genes at the expense of others.

Hamilton's rule is simple, but is it true? There is a whole literature (reviewed in more detail elsewhere³) that tries to answer this question. Queller's contribution is to provide a very simple and direct proof of Hamilton's rule, as an antidote to what I feel is the biologically unsympathetic and over-mathematical treatment that is usually applied. The positive feature of Queller's approach is that it sacrifices an element of mathematical rigour for the sake of deriving a much more general result, and deriving it with clarity and simplicity.

Complete recursion, or dynamic sufficiency, is the element of rigour sacrificed by Queller's approach. In a simple model (one locus, two alleles) in a diploid species, we can say that a rigorous model uses the frequencies of all three genotypes in one generation to predict all three genotype frequencies in the next. Complete recursion is the use of those genotype frequencies for predicting the genotype frequencies in the following generation, and so on. Queller's method has an incomplete recursion, because it uses all three genotype frequencies in one generation to predict the frequencies of the two genes in the next. Without making some assumption

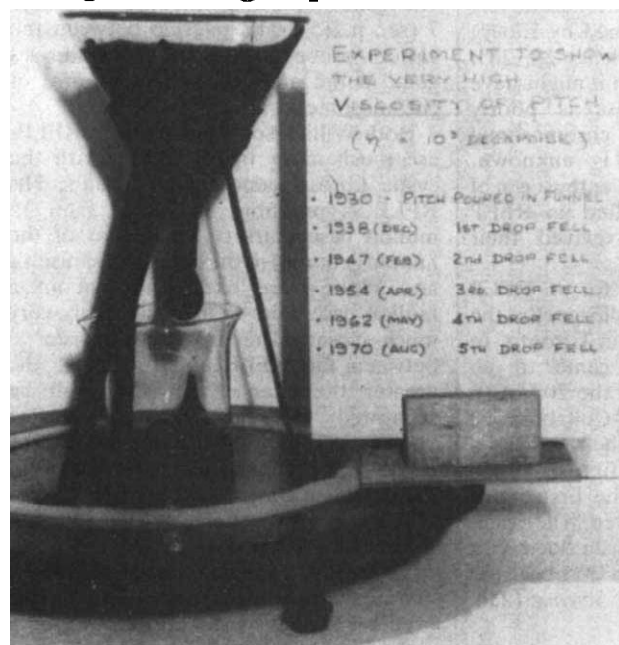
(usually that the Hardy-Weinberg equilibrium holds), genotype frequencies cannot be calculated from gene frequencies, and so it is impossible to apply Queller's equations repeatedly to predict gene frequencies in succeeding generations.

That is the cost of Queller's simplification. Whether it is serious depends on one's point of view. One view is that incomplete recursions are unsafe, because to predict gene frequencies into the future they must make some false assumption (such as Hardy-Weinberg), and since the consequences of the assumption may not be innocuous, the only way to provide a secure argument is not to make the assumption in the first place. But not all conclusions that we draw from evolutionary models depend on predicting gene frequencies in the future. Queller's model predicts whether a character increases or decreases from one generation to the next, and shows that it depends on a form of Hamilton's rule. He makes no claim about dynamic sufficiency (although he does unguardedly claim his model is "exact", a term usually reserved for completely recursive models), nor is any needed for his conclusion to be both interesting and important.

The advantages gained by forgoing complete recursion are great. Since Queller's model does not specify what causes the genetic similarity between donor and recipient, his result holds whether the cause is a recent common ancestry, heterogeneity of the population, or the tendency of individuals to interact preferentially with like genotypes for other reasons. His interactants need not comprise mutually exclusive and exhaustive groups. Completely recursive models are hard work: they assume (contingently, not necessarily) mutually exclusive and exhaustive groups, and they must specify exactly how genetic similarity is caused. For example, a model in which each group comprises half siblings would be quite different from one in which each group was formed by picking individuals at random from two sets of full siblings. Queller's approach shows that the relatedness is one quarter in each case, and so the two are immediately seen to be essentially similar. Completely recursive models would be very different and — though for no very clear reason — would eventually give the same answers as each other (and as Queller's model) in the matters important for the evolution of a character. The models would differ in details of interest to the population geneticist; for instance, they would have different kinds of polymorphic equilibria when the two alleles were so similar in effect as to be empirically indistinguishable.

While, in my view, Queller's trade-off leaves him as a champion of Hamilton's rule in one battle, he has been too ready to concede defeat on its behalf in another. Hamilton's rule was first derived for additive social interactions, in which the

Long-running experiments (I)



Only six drops have fallen, the latest in April 1979, since this experiment was set up in 1930. Its aim is to demonstrate and measure the fluidity and very high viscosity of pitch. The experiment, which continues in the Department of Physics at the University of Queensland, Australia, is one of three long-standing experiments considered in detail in the *European Journal of Physics* 5, 193–200 (1984), along with a request for accounts of further examples.

effects of others on one individual's number of offspring simply added up. What if interactions combine synergistically, as Queller calls it? The problem here is one of measurement. The third, synergistic, term in Queller's form can be made to disappear by agreeing to define benefit and cost as the average effects on individual's fitnesses, rather than as arbitrary terms in a model of fitness³.

So in Queller's simple model, Hamilton's rule, with costs and benefits correctly understood, is perfectly adequate for deciding the direction of change in gene frequency. In more complex models, in which individuals can pay costs and receive benefit many times in a lifetime, a real problem of non-additivity can arise. But one way round this is to make the plausible assumption that the gene effects in question are small; Fisher's microscope argument⁴ — that near to a good design only small changes are likely to be advantageous — suggests that, in the final stages of the perfection of a character, only genes of small effect matter.

Thus, for genes of small effect, additivity is restored and the correctness of Hamilton's rule is restored with it⁵. (Population geneticists may be interested in genes of large effect, and it is perfectly reasonable to doubt if characters are anywhere near perfection, so not everyone will be happy with this assumption. Note that a gene can have a small effect through rarity of expression even if it has a large effect on the occasions it is expressed.)

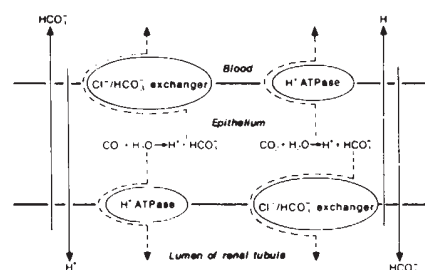
The method used by Queller is Price's covariance selection mathematics, a powerful technique of general value⁶ that was originated and expounded by Price in brief articles^{7,8}. For Queller, the value of

the technique was that it enabled him to change easily from counting genes in offspring according to the individual producing the offspring (the usual population genetics method of accounting by results), to counting according to the individual whose genotype caused the existence of the offspring (the inclusive fitness method of accounting by control).

In closing, I must mention a paper unjustly neglected by Queller, in which Price's method was first applied to inclusive fitness problems. This remains the most general and satisfying derivation and justification of inclusive fitness, at the same time as being remarkably simple. Queller cannot appeal to the obscurity of the journal in which the paper appeared, but he can plead for clemency on the grounds that its author himself later neglected the existence of his own strong and powerful result. Readers may have guessed that the author in question is W.D. Hamilton⁹. His hitherto unsung primacy in providing a simple and rigorous justification for inclusive fitness, together with his well-recognized provision of the idea in the first place, inspire me to end close to where I began: Hamilton rules OK! □

1. Hamilton, W.D. *J. theor. Biol.* **7**, 1 (1964).
2. Queller, D.C. *Nature* **318**, 366 (1985).
3. Grafen, A. *Oxford Surveys in Evolutionary Biology* **2**, 28 (1985).
4. Fisher, R.A. *The Genetical Theory of Natural Selection* (Dover Publications, New York, 1985).
5. Price, G.R. *Nature* **227**, 520 (1970).
6. Price, G.R. *Ann. hum. Genet.* **35**, 485 (1972).
7. Hamilton, W.D. *Nature* **228**, 1218 (1970).
8. Michod, R.E. & Hamilton W.D. *Nature* **288**, 694 (1980).

Alan Grafen is a Royal Society 1983 University Research Fellow at the Zoology Department, Oxford University, Oxford OX1 3PS, UK.



Depending on the respective locations of an ATPase and an exchanger, controlled by the acid load, the intercalated cell of the cortical collecting tubule either secretes H⁺ (left) or HCO₃⁻ (right).

of tubule after loading, compared with 129 per mm before loading. But the number of acid-secreting cells increases 10-fold, from 7 to 74, and the number of HCO₃⁻-secreting cells decreases correspondingly, from 122 to 63. This is evidence for a change in direction of transport in the same cell, rather than proliferation of a cell type with a certain direction of transport.

The change of direction is a result of the simple reversal of the arrangement of the transport mechanism, which consists of a proton-pump ATPase at one cell face and a Cl⁻/HCO₃⁻ exchanger at the opposite face (see figure). In H⁺-secreting cells the ATPase is at the luminal membrane and the exchanger at the blood-facing membrane; these locations are reversed in HCO₃⁻-secreting cells. The ATPase originates from intracellular vesicles which fuse with either the luminal or the blood-facing membrane, so that the ATPase is inserted into the appropriate membrane. Thus, one critical link in the control of transport direction by acid load is control of direction of vesicle movement.

These experiments open the way to exploring other steps in the chain of controls in the CCT. Insertion of the Cl⁻/HCO₃⁻ exchanger has not been studied and the detailed time course of acid control of ATPase insertion is not yet known. But this study also has broader significance for the developmental biology of epithelia in general. Biologists have long wondered how the genetic information in a radially symmetrical egg becomes translated into radially asymmetrical cells, tissues and animals. In epithelia, this question is posed in a particularly simple form, as functions are segregated between two cell faces. Membrane biologists have learned much about the details of the transport mechanism at each face but have so far paid little attention to developmental questions of how each mechanism gets inserted into the 'right' end of the cell. Having a cell in which the direction of insertion can be experimentally reversed may make it easier to answer questions such as these. □

Jared M. Diamond is Professor of Physiology at the University of California Medical School, Los Angeles, California 90024, USA.

Developmental biology

A reversible epithelium

from Jared M. Diamond

EPITHELIA are the sheets of cells that line body cavities such as the intestinal lumen. They are responsible for active transport of specific solutes, either from lumen to blood or vice versa depending on the particular solute and epithelium. In most cases the direction of transport is fixed, even though the rate of transport may be subject to physiological regulation. On page 368 of this issue G.J. Schwartz, J. Barasch and Q. Al-Awqati describe an unusual case of a reversible epithelium, which can transport the same solute in either direction in response to physiological need. The whole arrangement of pumps, exchangers and vesicles within the epithelial cell is reversible in detail. The tissue not only is a curious violation of the usual rules but also offers a promising system for learning how epithelia become programmed during development.

Schwartz *et al.* have studied the inter-

calated cells of the cortical collecting tubule (CCT) in the kidney. These regulate blood pH by either secreting H⁺ and absorbing HCO₃⁻, or secreting HCO₃⁻ and absorbing H⁺. The main physiological stimulus converting the tubule from net HCO₃⁻ secretion to net H⁺ secretion is the acid load. Schwartz *et al.* identify H⁺-secreting or HCO₃⁻-secreting cells by the endocytosis of fluorescent markers from the luminal or blood surface of the tubule, respectively. Since epithelial cells are commonly found to transport a given solute in a fixed direction, one might expect the CCT to contain two H⁺/HCO₃⁻-transporting cell types with opposite functions and the mitotic proliferation (or transport rate) of each cell type to respond to pH levels. Instead, mitotic figures are absent in intercalated cells of acid-loaded animals and the total number of cells remains virtually unchanged — 137 per mm

Particle physics

How to detect solar neutrinos

from W. Hampel

EXPERIMENTS designed to observe the neutrinos produced in the fusion reactions in the interior of the Sun provide the only way to test the theories of energy generation in stars. Another motivation for solar-neutrino experiments comes from particle physics: they may be the only possible way to obtain information on neutrino mixing (oscillations), a question of great importance in Grand Unified Theories. This is because neutrino mixing could be observable only in experiments with very large source-detector distances. Two recent papers^{1,2} put forward a new concept for a solar-neutrino detector and evaluate its ability to detect neutrino oscillations.

Neutrino mixing occurs if neutrinos are superpositions of two or more mass eigenstates with different rest masses. This leads to oscillations of the electron neutrinos, ν_e , produced in the Sun with neutrinos of other flavours (muon- or tau-neutrinos, ν_μ or ν_τ) on their way from the Sun to the detector. In fact, the discrepancy between the theory and observation for the only solar-neutrino experiment carried out so far, the Brookhaven Chlorine Experiment³, may be due to neutrino oscillations. Whether this is true or not will be decided by the outcome of the Gallium Experiment⁴, the only other solar-neutrino detector that has had its feasibility demonstrated in a pilot experiment.

Since both chlorine and gallium detectors employ the radiochemical technique that is based on inverse β decay, they are sensitive only to ν_e and measure the integrated (above their respective energy thresholds) energy spectrum of solar neutrinos. Therefore, they give only partial information on neutrino oscillations (through a depleted rate compared to that expected without oscillations). To learn more about neutrino oscillations, it will be necessary to use a detector able to give direct information on the energy spectrum of solar neutrinos. The detector proposed by Cabrera *et al.*¹ has this feature. It is based on the elastic scattering of neutrinos from electrons. Figure 1a presents the spectrum of the proton-proton (pp), ^7Be and ^8B solar-neutrino sources (the other solar-neutrino sources are less important and have been omitted from this figure) along with the recoil electron spectra expected for these sources from the $\nu_e + e^-$ reaction. Solar-neutrino detectors based on this reaction have been proposed before (for example, refs 6 and 7). In these proposals, the detection of the recoil electrons is through their Cerenkov radiation in water. This method, however, is restricted to high energy (≥ 5 MeV) recoil electrons and thus only allows observation

of the ^8B neutrino component (see Fig. 1).

The detection method suggested by Cabrera *et al.* is completely different. It is based on observing the temperature change of a macroscopic silicon mass (of the order of 1 kg), when a recoil electron is produced through a neutrino interaction and deposits its energy in the silicon block. For instance, a 100 keV recoil electron would raise the temperature of 1 kg Si from 1 mK to about 4 mK. This principle,

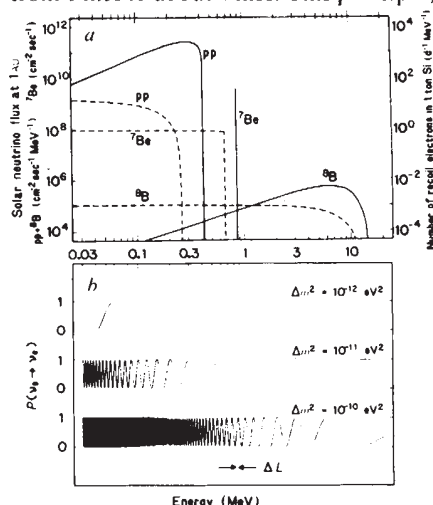


Fig. 1 a, Energy spectrum of the solar pp, ^7Be and ^8B neutrinos with the corresponding recoil electron spectra produced in the $\nu_e + e^-$ reaction. b, Probability $P(\nu_e \rightarrow \nu_e)$ of an electron neutrino ν_e emitted in the solar core reaching a terrestrial detector as ν_e . P as a function of neutrino energy is plotted for different Δm^2 .

the bolometric detection of solar neutrinos, was first investigated by Drukier and Stodolsky⁸ but their proposal was based on the coherent scattering of neutrinos from nuclei through the neutral current interaction and on a different technique (superconducting grain detector).

There are two major problems connected with bolometric detection of neutrinos. One is the problem of measuring small temperature changes in the mK region. The other is how to reduce the background to the very low level required for the detection of neutrinos. Concerning the temperature measurement, Cabrera *et al.* suggest, for instance, using superconducting thin film, such as tungsten deposited on the surface of the Si block. Determination of the resistance of this film would provide a measure of the temperature with a precision of about 10 keV in the energy range 10 keV–2 MeV.

Background is the more severe problem: the difficulties may be even larger with bolometric neutrino detection than with other types of detectors, since deposition of energy in the keV to MeV range is

not at all specific to neutrinos. All radiations, either from radioactive impurities in the Si itself or from outside sources (α -, β -, γ -radiation, neutrons, cosmic-ray particles), can dissipate energy in this interesting range and thus mimic a neutrino-scattering event.

As a first step, Cabrera *et al.* suggest constructing a prototype detector using a single Si block (~ 1 kg) and studying its properties with non-neutrino sources. Next, a larger detector (~ 100 kg) would be built for use in a reactor anti-neutrino experiment. Finally, a solar-neutrino experiment is planned with about 10 tons of Si ($\sim 10,000$ single cells).

Assuming such a detector could ever be built, what could be learnt from it about solar-neutrino oscillations? Restricting the discussion to the case where mixing of the ν_e occurs with only one neutrino of another flavour, the important quantity is the probability P that a ν_e emitted in the centre of the Sun is still (or again) a ν_e when it arrives at the detector:

$P(\nu_e \rightarrow \nu_e) = 1 - \sin^2 2\lambda \cdot \sin^2(\pi L \Delta m^2 / 2.48 E_e)$, where λ is a mixing angle ($0 < \lambda < \pi/4$), L is the distance between source and detector ($L = 1 \text{ AU} = 1.5 \times 10^{11} \text{ m}$), $\Delta m^2 = |m_1^2 - m_2^2|$ is the difference of the squared masses of the two involved neutrino mass eigenstates, in eV^2 , and E_e is the neutrino energy in MeV. $P(\nu_e \rightarrow \nu_e)$ as a function of the neutrino energy is shown in Fig. 1b for maximal mixing ($\lambda = \pi/4$) and three different values of Δm^2 . Thus, the original solar-neutrino spectrum Φ (Fig. 1a) is modulated as a function of energy with P yielding ΦP as the ν_e spectrum incident at the detector site.

Neutrino oscillations affect the recoil electron spectrum produced by the $\nu + e$ reaction because the cross-section for ν_e is about six times larger than that for neutrinos of other flavours. This is because the neutral current (exchange of Z^0 bosons) contributes to the cross-sections for neutrinos of all flavours, whereas the charged current (exchange of W bosons) affects the $\nu_e + e^-$ reaction only. Thus, neutrino oscillations result in reaction rates that are depleted compared with the situation without oscillations. An advantage is that this comparison can be carried out separately for the different solar-neutrino sources (in contrast to the case for radiochemical detectors). This is especially useful for the pp neutrinos, since their reaction rate in the absence of neutrino mixing can be predicted rather reliably without use of detailed solar models. Measurement of the depletion in the pp reaction rate thus yields P averaged over the pp-neutrino spectrum.

The most important quantity to be determined is Δm^2 . As can be seen from Fig. 1, different values of Δm^2 affect the ν_e spectrum incident at the detector in different ways. For the continuum ν sources, this in turn has an impact on the shape of the resulting recoil electron spectra. Thus, careful examination of these spectra

should reveal Δm^2 and, combined with the information on P , the mixing angle λ .

Among other interesting effects discussed by Krauss and Wilczek² (such as the modification of vacuum oscillations by matter effects) there are two concerning the nearly monoenergetic ^7Be neutrinos. The first occurs because the distance L between the Sun and Earth varies by 3.4 per cent in the course of the year due to the eccentricity of the Earth's orbit. Trivially, this leads to a $1/L^2$ variation of the total solar-neutrino flux at the Earth. Because the equation for P depends on L , this variation also causes a periodic horizontal displacement of the P -curves in Fig. 1b, with an amplitude indicated in Fig. 1 by ΔL . Thus, the ^7Be recoil electron rate is expected to vary in time with a period and an amplitude depending on both Δm^2 and λ , an effect observable for $\Delta m^2 > 10^{-10} \text{ eV}^2$ and values of λ that are not too small.

The second effect is not related to neutrino oscillations. Since the electrons captured by ^7Be nuclei in the solar core have a maxwellian velocity distribution depending on the local temperature T , the ν_e

emitted in this capture reaction have a spread in energy ($\sim 1 \text{ keV}$). Hence, the ^7Be recoil electron spectrum in Fig. 1 does not have a sharp cutoff at its endpoint energy. A measurement of the exact shape of this spectrum near the endpoint energy thus allows determination of T (requiring, of course, an energy resolution better than 10 keV).

Clearly, although the solar-neutrino detection suggested by Cabrera *et al.* and other detectors with similar capabilities in neutrino-energy spectroscopy are still far from being built, it would be well worthwhile investing in their development. $\square$

1. Cabrera, B., Krauss, L.M. & Wilczek, F. *Phys. Rev. Lett.* **55**, 25 (1985).
2. Krauss, L. & Wilczek, F. *Phys. Rev. Lett.* **55**, 122 (1985).
3. Rowley, J.K., Cleveland, B.T. & Davis, R. Jr *AIP Conf. Proc.* **126**, 1 (1985).
4. Hampel, W. *AIP Conf. Proc.* **126**, 162 (1985).
5. Hampel, W. *Proc. Neutrino 84*, 530 (World Scientific Publishing, Singapore, 1984).
6. Bakich, A.M. & Peak, L.S. *AIP Conf. Proc.* **126**, 238 (1985).
7. Chen, H.H. *AIP Conf. Proc.* **126**, 249 (1985).
8. Drukier, A. & Stodolsky, L. *Phys. Rev. D* **30**, 2295 (1984).

W. Hampel is at the Max-Planck-Institut für Kernphysik, 69 Heidelberg 1, FRG.

Vertebrate palaeontology

First marsupial fossil from Asia

from Michael J. Benton

MARSUPIALS, pouched mammals like kangaroos and opossums, are today restricted to Australia and the Americas (Fig. 1). How this split distribution came about has been much debated by biogeographers. A northern dispersal route was favoured at first, with the early marsupials travelling from the Americas across Asia to Australia, but no evidence of marsupials had been found at that time in Asia. The acceptance of continental drift led most people to prefer a southern dispersal route from South America to Australia via Antarctica (Fig. 2). Now that evidence of a fossil marsupial has

been reported from central asiatic USSR, the question is again open.

The specimen is a single tooth, identified as a first or second upper molar from the right side of the jaw. It is very small (1.7 mm long by 1.6 mm across) and its closest affinities seem to be with the opossums (family Didelphidae, subfamily Didelphinae). It is most like a tooth of the European fossil opossums *Amphiperatherium* and *Peratherium*, and the North American *Herpotherium*. The new Russian find comes from the middle of the Aksyr Formation, which is lowermost Oligocene (almost 37 Myr ago), in

the Zaysan territory in eastern Kazakhstan, central Asia.

The fossil record of marsupials was fairly sparse until a few years ago. The oldest forms came from the late Cretaceous (84–65 Myr) deposits in North America, in which fossils are quite common. Up to 30 species of early marsupials have been identified in Alberta, Montana and Wyoming; these are divided into three families, two of which died out with the dinosaurs at the end of the Cretaceous period and the Didelphidae (opossums), which has survived to the present day. A few late Cretaceous marsupials have also been reported from Peru and Bolivia in South America.

In the subsequent Tertiary period, the didelphid marsupials spread to Europe, where they survived until the Miocene (about 20 Myr). In North America, the didelphids also become extinct in the Miocene but then reinvaded from South America about three million years ago. During the Tertiary, many new groups of marsupials arose in South America, including equivalents of sabre-toothed cats and some dog-like forms, but only three families still survive.

There are now 13 families of marsupials in Australia. The fossil record is sparse, with only a few odd teeth and jaws known from the late Oligocene, but there are more abundant and diverse remains from the Miocene onwards.

Three years ago, a South American marsupial (a polydolopid) was found in the Eocene deposits (about 40 Myr) of Antarctica, a find that seemed to support the theory of a southern dispersal route, with an ice-free Antarctica as the bridge between the southern tip of South America and Australia. Then, two years ago, marsupial remains in Africa were reported by two independent teams^{1,4}—but these are probably the remains of animals that wandered into Africa from Europe and thus do not help to disentangle the palaeobiogeography of early marsupial dispersal.

Although the Asian tooth may seem to provide support for the northern dispersal theory, it is approximately equivalent in age to the oldest known Australian marsupials, and so occurs too late to fit with the theory. It is also a didelphid, a European–American form, with no particular Australian affinities. Most probably the didelphids were at one time distributed more widely than has been thought and then died out without leaving any descendants in Europe, Africa and Asia. $\square$

1. Gabunia, L.K. & Shevyreva, N.S. *Doklady Akad. Nauk, S.S.S.R.* **281**, 684 (1985) (translated by J. Wilson, Ulster Museum, Belfast).
2. Woodbourne, M.O. & Zinmeister, W.J. *Science* **218**, 284 (1982).
3. Mahboubi, M., Ameur, R., Crochet, J.-Y. & Jaeger, J.-J. *C. r. Acad. Sci., Paris* **297** (III), 415; *ibid.* 691 (1983).
4. Bown, T.M. & Simons, E.L. *Nature* **308**, 447 (1984).
5. Marshall, L.L. in *Aspects of Vertebrate History* (ed. Jacob, L.L.) 345 (Museum of Northern Arizona, Flagstaff, 1980).

Michael J. Benton is in the Department of Geology, The Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland.

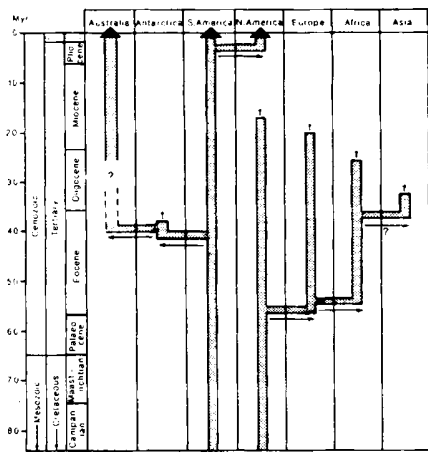


Fig. 1 Probable relationships between marsupial groups in different continents.



Fig. 2 The two possible dispersal routes with the positions of the continents as they were 80 Myr ago (modified from ref. 5).

Metamorphic geology

Tectonics of metamorphism

from Michael Brown

LARGE tracts of the continents of the Earth are formed by rocks which have undergone regional metamorphism — that is, their mineral assemblages indicate that they have been subjected to elevated temperatures and pressures at some time in their past. Typically, these areas are located in orogenic belts or Precambrian shields, and show an intimate relationship between metamorphism and tectonic deformation. Generally accepted tectonic settings for regional metamorphism are continental margins associated with subduction, such as the west coast of South America; island arcs, such as those of South-east Asia; and zones of continental collision, such as the Alps and Himalayas. But the Hercynian metamorphism of the Pyrenees does not fit easily into this framework, as it must be explained by a tectonic setting involving neither subduction nor continental collision. On page 330 of this issue¹, S.M. Wickham and E.R. Oxburgh use data from the Pyrenees to formulate a model of continental rifting, with or without strike-slip movement, as the setting for regional high-temperature metamorphism.

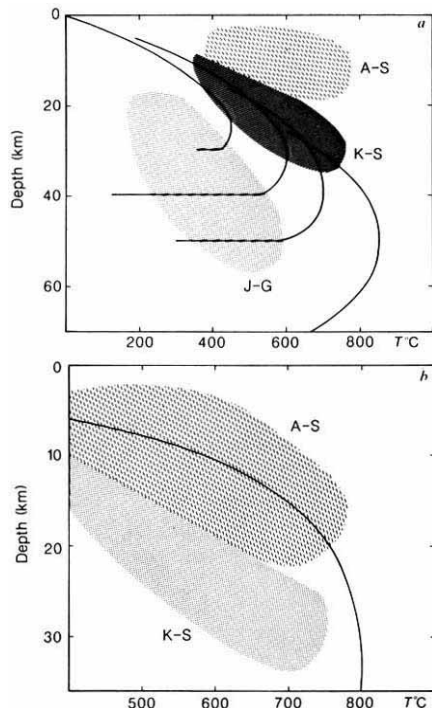


Fig. 1 *a*, P - T paths followed by rocks buried initially at 70, 50, 40 and 30 km after overthrusting (after Fig. 6c of ref. 6). A-S, andalusite-sillimanite facies series; K-S, kyanite-sillimanite facies series; and, J-G, jadeite-glaucophane facies series. *b*, Uplift P - T path followed by rocks initially metamorphosed at 800°C and at 35 km depth undergoing approximately adiabatic decompression; rate of uplift 2 mm per year (after model B of Fig. 9 of ref. 15). A-S and K-S as in *a*.

Metamorphic terrains commonly develop as paired belts, in which one member is characterized by blueschists and eclogites, which indicate low temperature and high pressure conditions and which is interpreted as having been developed at a site of low heat flow, such as a subduction zone; the other member is characterized by high-temperature metamorphism and the development of migmatites and anatectic granites, and is interpreted as having been developed at a site of high heat flow, such as that beneath an associated volcanic arc. In detail, regional metamorphic terrains have been classified by their mineral assemblages into facies series types such as andalusite-sillimanite, kyanite-sillimanite and jadeite-glaucophane. It is implicit in this classification that an entire metamorphic terrain can be described by a single geothermal gradient and that higher-grade mineral assemblages develop from lower grade ones similar to those now found along the present erosion surface. This traditional view has been replaced by one in which individual rocks³ and minerals⁴ can be used to derive paths in pressure-temperature (P - T) space, which can be related to the tectonic setting^{5,6}. The loci of peak or slightly post-peak P - T conditions as preserved by the mineral assemblages — the metamorphic geotherm — results from the intersection of P - T paths for individual rocks with the erosion surface. It is this which is represented by the particular facies series.

Assuming that a uniformitarian approach to the problem is valid, we can try to unravel the tectonic history of metamorphic terrains by relating the characteristic features of young metamorphic belts to their tectonic settings. In essence, this means establishing the characteristic P - T paths of metamorphism in different plate-tectonic settings. For instance, the jadeite-glaucophane facies series can be generated by subduction. However, the Pyrenees comprise a series of structural and thermal domes with condensed metamorphic zonal sequences of the andalusite-sillimanite facies series, culminating in migmatites (rocks that have been partially melted) and abnormally high crustal temperatures and hence thermal gradients. It is this type of metamorphism that Wickham and Oxburgh explained.

Pioneering work on metamorphism associated with continental rifting⁷ and metamorphism due to magmatic accretion⁸ did not lead to a level of understanding of the thermal structure associated with such metamorphism comparable to our knowledge of overthrust terrains^{2,5,6}, although both are potentially important in

the island arc and active continental margin settings which so dominate lithospheric plate margin interactions. It is in this latter context that the work of Wickham and Oxburgh¹ is important.

Although it has been developed for a metamorphic terrain that cannot have formed at a subducting margin, the model has implications for high-temperature metamorphism in general, as extension and strike-slip movements are common in island arcs and active continental margins⁹. However, Precambrian meta-



Fig. 2 Typical texture produced by overstepping of reactions during approximately adiabatic decompression. In this case, garnet (black) plus quartz (medium grey in centre of view) reacts to orthopyroxene (granular mineral surrounding quartz crystal) plus cordierite (white to grey mineral in central part of view clearly replacing garnet); garnet-orthopyroxene-cordierite gneiss from the Sharyzhalgay Complex, USSR. Long dimension of photomicrograph is 5.5 mm, crossed polarized light.

morphic terrains such as the Napier Complex of Antarctica and the Arunta Complex of Australia, which have suffered extreme temperatures (900–950°C) and are characterized by approximately isobaric cooling to more moderate temperatures (600–650°C) during substantial periods of time, do not fit into a uniformitarian model of metamorphism based on plate tectonics theory. Moreover, the role of magma in regional metamorphism is poorly understood, although examples have been described from Greenland¹⁰ and Australia¹¹.

During continental collision, such as in the Alps, regional metamorphism may be related to crustal thickening, either by overthrusting or by uniform strain. The result is a characteristic anticlockwise loop in depth-temperature space (Fig. 1*a*), with one third or more of the P - T path being within 50°C of the maximum temperature achieved during burial and uplift¹², and high-pressure metamorphism followed by increasing temperature and decreasing pressure.

More difficult to relate to tectonic setting with certainty are those terrains characterized by overstepping of reactions during decompression^{12–14} (Figure 2 illustrates an intermediate stage of a reaction). It is in the context of these metamorphic terrains characterized by approximately adiabatic decompression (Fig. 1*b*), that

the model for high-temperature metamorphism proposed by Wickham and Oxburgh¹ for the Pyrenees is important. Other essential features of the Pyrenean belt are the contemporaneity of sedimentation and metamorphism and the post-metamorphic development of granitoids in the lower crust. Such features are common in other high-temperature metamorphic belts such as British Columbia² and southern Brittany³, as well as in those found in Japan. The suggestion of Wickman and Oxburgh that the high-temperature member of paired metamorphic belts might develop by rifting and extension in island arcs is plausible and attractive. It is also timely, coming before a Royal Society meeting (January 1986) on Tectonic Settings of Regional Metamorphism. □

1. Wickham, S.M. & Oxburgh, E.R. *Nature* **318**, 330 (1985).
2. Oxburgh, E.R. & Turcotte, D.L. *Schweiz. miner. petrogr. Mitt.* **54**, 641 (1974).
3. Thompson, A.B. & England, P.C. *J. Petrol.* **25**, 929 (1984).
4. Spear, F.S. & Selverstone, J. *Contrib. Mineral. Petrol.* **83**, 348 (1983).
5. England, P.C. & Richardson, S.W. *J. geol. Soc. London* **134**, 201 (1977).
6. England, P.C. & Thompson, A.B. *J. Petrol.* **25**, 894 (1984).
7. Falvey, D.A. *APEA Journal* **14**, 95 (1974).
8. Wells, P.R.A. *Earth planet. Sci. Lett.* **46**, 253 (1980).
9. Tamaki, K. *Geology* **13**, 475 (1985).
10. Wells, P.R.A. *J. Petrol.* **20**, 187 (1979).
11. Vernon, R.H. *Geology* **10**, 76 (1982).
12. Hollister, L.S. *Can. Mineral.* **20**, 319 (1982).
13. Brown, M. *Migmatites, Melting and Metamorphism* (ed. Atherton, M.P. & Gribble, C.D.) 174 (Shiva Publishing, Nantwich, 1983).
14. Harris, N.B.W. & Holland, T.J.B. *Am. Mineral.* **69**, 1036 (1984).
15. Albarède, F. *Soc. Geol. France Bull.* **18**, 1023 (1976).

Michael Brown is Head of Geological Sciences at Kingston Polytechnic, Penrhyn Road, Kingston upon Thames, KT1 2EE, UK.

Human genetics

Measuring mutation in man

from Mary F. Lyon

GENETIC disease is an important cause of ill health and of physical and mental handicap, particularly in childhood, and results in a significant number of premature deaths. Major advances in preventing the birth of genetically affected children are being made, through genetic counselling, prenatal diagnosis and the use of recombinant DNA methods. But mutation remains a problem, resulting in the levels of many genetic diseases being maintained despite adverse selection, and it is becoming increasingly important to assess the possible increase in mutation rates caused by chemical or physical agents in the environment. At two recent meetings* plans were put forward for an international collaboration to measure the mutation rates in man by studying the offspring of cancer patients who have survived to reproduce after radiotherapy and chemotherapy. This should, for the first time, provide information on the effect of increased mutation rates on the incidence of hereditary diseases in man.

Genetic diseases may cause death before reproductive age, like Duchenne's muscular dystrophy, or result in failure to reproduce through, for example, severe mental defect due to fragile-X syndrome. Such diseases are maintained at their present frequency by a balance between fresh mutation and adverse selection. New mutations mean that affected individuals occur without any previous history of the disease in a family, making prenatal diagnosis of limited usefulness. Intensive efforts to clone the gene for Duchenne's muscular dystrophy were discussed recently in *Nature*^{1,2} and, when successful, will make it possible to prevent the birth of

affected sons to known carrier mothers. But one third of all cases are due to new mutation and the means of preventing these is unclear³, as it would be impractical to screen the whole population.

If the mutation rate rises, the incidence of such mutationally maintained diseases will rise, so efforts are being directed towards the detection of mutagens in the environment and their control or elimination. The International Commission for Protection against Environmental Mutagens and Carcinogens (ICPEMC) aims to evaluate scientific data relating to environmental mutagens and carcinogens and to provide information for estab-

lishing guidelines for regulatory or legislative action.

The question of detection and protection against germ-cell mutagens in man is highly complex. So far, successful methods of detecting changes in mutation rate directly in human populations have not been found⁴. Extrapolation from experiments is thus necessary, but this is only possible with data obtained in mammalian germ cells, because there are wide variations in the sensitivity of different cell types, which are not understood⁵. Germ-cell data are laborious and costly to obtain and so are not available for many agents.

Relatively extensive information for several cytotoxic drugs that are used in cancer chemotherapy or as immunosuppressants does, however, indicate that some of these agents, for example cyclophosphamide⁶, are strongly mutagenic for certain germ-cell stages. With the recent improvements in the treatment of cancer occurring in children and young adults, many survivors of cancer treatment are living to reproduce. For the sake of the patients and their children there is a clinical need to know the long-term prospects, not only about the likelihood of second cancers in the patients but also about their fertility and the normality of their children.

Investigations into the reproductive histories of surviving cancer patients have already begun and preliminary results were presented at the first meeting. Examples include surveys of childhood cancer in Italy (G. Pastore, Turin), the United Kingdom (G. Draper, Oxford) and the United States (J. Mulvihill, Bethesda); of testicular cancer in Norway (S. Fossa, Oslo) and the United Kingdom; and of Hodgkin's disease in the United States (S. Donaldson, Stanford). In the

100 years ago



The transit instrument of the Bedford Observatory, with observing chair and clock. From *Nature* **33** 59, 19 November 1885.

*Workshop on possible genetic effects in offspring of surviving cancer patients, Oslo, May 3-5, 1985 (sponsored by ICPEMC and the Norwegian Cancer Society). Follow-up meeting, Lyon, 1-3 October, 1985 (sponsored by IARC).

course of the meeting, it became clear that it would be very difficult for surveys in a single country to recruit sufficient patients and offspring to obtain meaningful results. Extrapolations from animal data suggest that many thousand offspring of patients treated with cytotoxic drugs will have to be studied to detect an increase in genetic disease resulting from such treatment⁶. Examination of survivors from atomic bomb explosions have also shown that large numbers are needed to detect the effects of radiation⁴.

At the first meeting, therefore, an international collaborative study was suggested, using a standardized protocol so that results could be pooled. A smaller group, including representatives of the Union Internationale contre Cancer, later met to formulate practical arrangements for initiating the study. The types of cancers to be studied would include childhood cancers, testicular tumours, and Hodgkin's and other lymphomas. Transplant patients and others treated with immunosuppressive drugs known to be mutagenic, such as cyclophosphamide^{7,8}, might also be included. Records would be kept of fertility, outcome of pregnancy including spontaneous abortions, birth weight, malformations and cancer in the offspring. Samples of blood would also be obtained from the patient, the other parent and the offspring, and stored in the hope that studies of chromosomes, protein (K. Berg, Oslo) or DNA variants would prove possi-

ble — potentially these could yield much more detailed information than records of clinical abnormalities⁴. At present, however, such studies are much more laborious and expensive; the DNA studies, though potentially the most informative, are also the most expensive (P. Pearson, Leiden).

Such a programme would benefit the patients and their offspring by providing information on the levels of genetic risk incurred from ever more aggressive cancer treatments. It is possible that the relative mutagenicity of different treatments could be assessed, so that the genetic risk could be reduced. Scientifically, the results could be compared with the extrapolations already made from animal work⁶ to assess the validity of such extrapolation. Most important, data would become available on the effects that known doses of strong mutagens have on germ-cell mutation in man. This will be a major step in tackling the problem of mutation as a cause of human genetic disease. □

1. Monaco, A. P. *et al.* *Nature* **316**, 842 (1985).
2. Connolly, P. M. *Nature News and Views* **316**, 763 (1985).
3. Kedes, L. H. *Trends Genet.* **1**, 205 (1985).
4. Schull, W. J., Otake, M. & Neel, J. V. *Science* **213**, 1220 (1981).
5. Neel, J. V. *J. Hered.* **74**, 2 (1983).
6. Lyon, M. F. *et al.* *Mutat. Res.* **115**, 225 (1983).
7. Brown, N. A. *Nature News and Views* **316**, 110 (1985).
8. Trasler, J., Hales, B. & Robaire, B. *Nature* **316**, 144 (1985).

Mary F. Lyon is in the Medical Research Council Radiobiology Unit, Chilton, Didcot, Oxon OX11 0RD, UK.

models for the epochs 1945.0, 1950.0, 1955.0 and 1960.0 (IGRF 1945, IGRF 1950, IGRF 1955 and IGRF 1960).

DGRF now spans the interval 1965.0 to 1980.0 with four main-field models for 1965.0, 1970.0, 1975.0 and 1980.0 (DGRF 1965, etc.). For dates between the epochs of the models, linear interpolation between the coefficients of the two models on either side of the date is to be used. A similar procedure is to be used for dates in the interval 1945.0 to 1965.0, using the IGRF 1945, IGRF 1950, IGRF 1955, IGRF 1960 and DGRF 1965 models, as appropriate. Extrapolation back to 1940.0 will probably be reasonably accurate, though this was not formally recommended by the working group.

Further revision of DGRF is not anticipated. The pre-1965 models (IGRF 1945 to IGRF 1960) will probably be replaced by definitive models in 1987. The newly adopted DGRF 1980 model replaces the former PGRF 1975 and IGRF 1980. The present PGRF 1980 will be superseded when a definitive model of the main-field at 1985.0, different from IGRF 1985, is adopted.

The main-field models for 1960 to 1985 have 120 coefficients each and extend to degree and order 10. The main-field models for 1945 to 1955 and the predictive secular-variation model have 80 coefficients and extend to degree and order 8. The coefficients are given in the Schmidt quasi-normalized form¹ and refer to a sphere of radius 6,371.2 km. When converting between geodetic and geocentric coordinates the use of the international atomic unit (IAU) ellipsoid² is recommended; this ellipsoid has an equatorial radius of 6,378.16 km and a flattening of 1/298.25.

The coefficients of the IGRF models and computer programs for synthesizing field values are available from:

- World Digital Data Centre C1, British Geological Survey, Murchison House, West Mains Road, Edinburgh EH9 3LA, UK.
- World Data Center A, National Oceanic and Atmospheric Administration, EDIS/NGSDC (D62), 325 Broadway, Boulder, Colorado 80303, USA.
- World Data Center A for Rockets and Satellites, Code 601, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. □

1. International Geomagnetic Reference Field 1965.0. *J. geophys. Res.* **74**, 4407 (1969).
2. International Geomagnetic Reference Field 1975. *J. geomagn. Geoelect.* **27**, 437 (1975).
3. International Geomagnetic Reference Fields: DGRF 1965, DGRF 1970, DGRF 1975 and IGRF 1980. *EOS Trans. Am. geophys. Un.* **57**, 120 (1981).
4. Zmuda, A. J. *Bull. Int. Ass. Geomagn. Aeronom.* **28**, 148 (1971).
5. Peddie, N. W. *J. geomagn. Geoelect.* **34**, 309 (1982).
6. Chapman, S. & Bartels, J. *Geomagnetism*, 611 (Clarendon, Oxford, 1940).
7. International Astronomical Union *Proc. 12th General Assembly* **12B**, 594 (1966).

David R. Barraclough is at the British Geological Survey, West Mains Road, Edinburgh EH9 3LA, UK.

Geophysics

International Geomagnetic Reference Field revision

from David R. Barraclough

THE International Geomagnetic Reference Field (IGRF) is a series of mathematical models of the main geomagnetic field and its secular variation. The models consist of sets of spherical harmonic (or Gauss) coefficients. IGRF has become widely used for deriving values of geomagnetic components used in, for example, studies of magnetic anomalies and investigations of charged particle motions in the ionosphere and magnetosphere.

Since it was first adopted by the International Association of Geomagnetism and Aeronomy (IAGA) in 1968 (IGRF 1965, ref. 1), IGRF has been revised three times: IGRF 1975 (ref. 2); IGRF 1980 (ref. 3); and now IGRF 1985. Details of the derivation of the original IGRF and of its development up to 1981 have been given by Zmuda⁴ and Peddie⁵.

The latest revision of IGRF was considered by Working Group 1* (analysis

of the main field and secular variations) of IAGA Division I during the Fifth General Assembly of IAGA held in Prague in August 1985. The following additions and modifications to IGRF 1980 were recommended.

- The extension of the definitive international geomagnetic reference field (DGRF) to 1980.0 by the adoption of a new model (DGRF 1980) to replace IGRF 1980.
- The addition of an international geomagnetic reference field for the interval 1985.0 to 1990.0 (IGRF 1985) consisting of a model of the main field at 1985.0 and a predictive model of the secular variation for use in adjusting the main-field model to dates between 1985.0 and 1990.0.
- The adoption of a provisional international geomagnetic reference field for the interval 1980.0 to 1985.0 (PGRF 1980), defined by linear interpolation between the coefficients of DGRF 1980 and IGRF 1985 (main field).
- The addition of a series of main-field

*The membership of Working Group 1-1 is: D. R. Barraclough (chairman), W. Mundi (vice-chairman), F. S. Barker, V. P. Golovkov, P. J. Hood, F. J. Lowes, N. W. Peddie, O. Guizhong, S. P. Srivastava, R. Whitworth, D. E. Winch, T. Yukutake and D. P. Zidarov.

Meteorite impacts on humans and on buildings

SIR—From nine years of observation with the Meteorite Observation and Recovery Project, a network of 60 cameras in western Canada, we have derived the frequency of meteorite falls on the Earth as a function of the total mass of meteorites to reach the ground for each event¹. Assuming that the total mass of an event is twice the mass of the largest fragment observed by the cameras, we conclude that

$$\log N = -0.689 \log m + 2.967$$

where N is the number of events per year in which a mass of at least m grams of meteorite is deposited in an area of 10^6 km². This formula predicts 39 events per year with a mass of at least 100g in each million km², or 5,800 such events on the total land area of the Earth.

Our present concern is to estimate the probability of persons or buildings being struck by meteorites. Our major assumptions are: (1) A human being occupies an area of 0.2 m². (2) The smallest impact likely to be reported would involve a fragment of a few grams. (3) Typical meteorite falls consist of five major fragments. (4) Meteorite fragments larger than 200 g will normally penetrate a roof and ceiling. (We also assume that, if the total mass exceeds 500 g, each of the five fragments could penetrate a roof, whereas none of the pieces from lesser events will do so.) (5) Residents of North America spend 5 per cent of each day outside and 95 per cent of the day protected by a roof. (6) The total roof area of buildings averages 50 m² per member of population.

On the basis of these plausible assumptions we derive the frequency of impacts on people and on buildings and we wish to compare the values with recent experience. For this purpose we consider the population of the United States plus Canada (about 2.5×10^8 persons) since we expect news coverage of relatively minor events to be more complete for this sample than in many densely populated areas of the world. For this North American sample, our assumptions predict an annual rate of 0.0055 impacts on people (one event per 180 years) and 0.80 impacts per year causing damage to buildings.

The only documented case of a person being struck by a meteorite appears to be a fall on 30 November 1954². A fragment of a stony meteorite weighing 3.9 kg penetrated the roof and ceiling of a house in Sylacauga, Alabama, bounced off a large radio and struck a woman who was asleep on a couch, inflicting painful bruises. At first glance it would appear unlikely that there would be even one known event only 31 years ago, but the fact that there are no other verified cases elsewhere in the world indicates that impacts on people are extremely rare.

The prediction for impacts on buildings is more readily verified because the pre-

dicted rate is much higher. All known meteorite falls and chance recoveries are reported in the *Meteoritical Bulletin*, published about once per year in the journal *Meteoritics*. During the past 20 years, there are reports of 16 recoveries from fresh meteorite falls in the United States and Canada. For seven of these events, there was appreciable damage to some building, usually the roof of a home or garage. In the Louisville, Kentucky fall in January 1977, three separate buildings were struck, so there are nine reports of damage. Two other events, which we have not included, involved a small meteorite which caused no damage to a roof and a 1.3 kg object that damaged a mailbox. It is clear that impact on a building greatly increases the probability of recovery of the meteorite, since only a very small fraction of all falls will strike buildings, whereas half of the recent recoveries have such an involvement.

We would predict 16 damaged buildings in 20 years for this sample, which is close to the nine reports of damage. Since we would not expect small meteorites to be located, identified and reported in all cases, especially for minor damage to commercial buildings where there might be little interest in finding the cause of a leaky roof, we suppose the actual rate may be somewhat higher than our estimate. Our guess of five effective fragments per event may be too low.

To extrapolate to the entire world population, estimated at 5,000 million, then the number of events will increase by a factor of 20 over the North American sample if we use the same assumptions. One would then expect a person to be struck by a meteorite once in nine years and that sixteen buildings per year would receive some damage from meteorites.

We thank Dr Paul A. Feldman of the Herzberg Institute for valuable discussions.

I. HALLIDAY
A. T. BLACKWELL
A. A. GRIFFIN

*Herzberg Institute of Astrophysics,
National Research Council of Canada,
Ottawa, Canada K1A 0R6*

1. Halliday, I., Blackwell, A. T., Griffin, A. A. *Science* **233**, 1405–1407 (1984).

2. Swindel, G. W. Jr & Jones, W. B. *Meteoritics*, **1**, 125–132 (1954).

A question of cellular immortality

SIR—In “Human T-lymphotropic retroviruses” (*Nature* **316**, 395–403; 1985), as well as other articles many authors now speak of the “immortalization” of cells by the expression of oncogenes. This term is generally used to refer to the ability of a cell line to grow *in vitro* indefinitely, without senescence or a growth crisis. The use of the term, to those not versed in the minutiae of oncogene nomenclature, would seem to imply specific gene(s) whose func-

tion is to prevent cell senescence. This gene would seem to have no corresponding physiological function *in vivo*, however.

A more plausible explanation would be that the cell, at a certain point in its step-wise differentiation, fails to turn off a regulatory gene. The cell continues to express this regulatory gene and the genes whose expression it controls. The cell is “stuck” in a state of expressing inappropriate genes and is prevented from terminal differentiation. This observation would fit with the fact that established cell lines, while not necessarily tumorigenic, require fewer steps to render them tumorigenic, perhaps as little as a growth factor with a specific effect on a certain differentiation state.

MARGARET PILLOW
SELINA BENDIX

*Bendix Environmental Research Inc.,
Fox Plaza, Suite 902,
1390 Market Street,
San Francisco, California 94102, USA*

Weak violation — a new concept in relativity?

SIR—Quoting from *Nature News and Views*, “If there are violations of special relativity, they are at most weak violations, departures from what is not strict orthodoxy and not flat contradictions of it”. Not everyone would agree with this statement. For many decades now, Einstein’s relativity, and the need for absolute conformity with its basic tenets, have stood firmly in the way of any theoretical advance which dares to imply the slightest violation. But though it is improbable that we will ever see support for Marinov’s claim, that the Earth’s cosmic motion through space can be detected by tests confined to the laboratory, should this be demonstrated as proven, then Einstein’s theory is violated by a death blow struck into its very heart.

Consistent with empirical data, a number of alternative laws of electrodynamics were shown by Clerk Maxwell to be derivable from a formula based on the relative velocity of interacting charges, taken in conjunction with what we would, today, term a “quantum electrodynamics” proposition (Fechner’s hypothesis). Einstein’s constraint, that any such law must satisfy the test of covariance, excludes a form of law which has recently been supported by experimental evidence from anomalous forces accompanying electric discharges in electrolytic solutions^{1,2}.

Consequently, research into practical applications of the indicated law has, at least until the present time, been retarded, blocked by the “inviolable” principle of relativity, and for the very reason that a successful outcome would be a strong violation and therefore improbable.

Whereas Marinov has sought to chal-

lenge Einstein directly on the speed of light issue, the Achilles' heel may already have been exposed in the application of electrodynamic law to electrochemical processes. Relativity, in its classical sense, will undoubtedly survive, but the real question is whether the test of time will continue to support Einstein's requirement that each and every relatively-moving observer constitutes his own personal reference frame for the locally-applicable laws of physics.

H. ASPDEN

Department of Electrical Engineering,
University of Southampton,
Southampton SO9 5NH, UK

1. Maddox, J. *Nature* **316**, 209 (1985).
2. Aspdén, H. *Phys. Lett.* **111A**, 22 (1985).
3. Graneau, P. & Graneau P.N. *Appl. Phys. Lett.* **46**, 468, (1985).

Structure and activity of barbiturates

SIR—It has long been a puzzle that caffeine and other purines should be stimulants while the chemically similar barbiturates are sedatives¹. I suggest that the latter are no longer planar in solution, being hydrated to the *gem*-diols and having the molecule folded about its principal axis (boat configuration).

Hydration of the central carbonyl can reasonably be expected because of the high positive charge on each of the neighbouring amide nitrogens. The result of electron-withdrawal near a carbonyl group is most familiar as chloral hydrate. Such *gem*-diols are known to catalyse the hydration of carbon dioxide, if the pH is high enough for partial ionization of the diol². Most barbiturates would then be fully ionized as amides, unless they were substituted at both nitrogens. Veronal, therefore, should be inactive, and etorbarbital active, in promoting the hydration of dissolved carbon dioxide to carbonate.

As for biological activity, *gem*-diols are ideally constituted to bind the ionized carboxyl of GABA and other amino acids.

T. NASH

Bower Chalke,
Salisbury, Wilts SP5 5BW, UK

1. Fieser, L.M. & Fieser, M. *Advanced Organic Chemistry*, 528 (Reinhold, New York, 1961).
2. Sharma, M.M. & Danckwerts, P.V. *Trans. Faraday Soc.* **59**, 386 (1963).

Linking impacts and plant extinctions

SIR—The discovery of shock-metamorphosed quartz grains¹, in addition to earlier evidence from siderophile enhancements² and microtektite-like spheroids^{3,4} appears to confirm the existence of a bolide event at the Cretaceous-Tertiary (K/T) boundary, but a causal connection between this event and the extinctions of many plants and animals remains in dispute. The pattern of extinctions in the marine realm⁵⁻¹⁰ generally conforms to

the hypothesized effects of an asteroid impact, but evidence from the terrestrial biota has been somewhat paradoxical. Though the situation for land animals remains equivocal (because some of their fossils are suspected to have been reworked), recent evidence collected from both fossil pollen and palaeosols appears to support the marine data for a sudden, catastrophic change at the K/T boundary.

Latest Cretaceous pollen sequences recovered from many sites in Montana^{11,12}, Colorado¹³ and New Mexico¹⁴ display dramatic shifts in the ratio of fern spores to angiosperm pollen at the K/T boundary. These widespread palynofloral disruptions suggest that the terrestrial flora suffered a profound ecological shock at the boundary, and the close proximity of these pollen breaks to Ir anomalies and highly shocked quartz grains implicate an asteroid impact. The pattern of plant destruction and renewal is similar to that observed at Tunguska¹⁵, Krakatau¹² and Mount St Helens¹⁶. Initial devastation might have been produced by the shockwave¹⁶ that would have resulted from a terrestrial impact¹⁷.

In addition, information collected from palaeosols (M.D.S. and G.J.R., unpublished data) in Montana tends to support the pollen evidence for a change in the flora at the K/T boundary. Palaeosols from the Hell Creek Formation are well developed, weakly calcareous and have a microstructure indicative of warm, oxidizing conditions. These Cretaceous palaeosols combined with pollen data are suggestive of a well drained closed-canopy forest of subtropical aspect, dominated by broadleaf angiosperms. Mean annual rainfall was subhumid, but periodic, possibly monsoonal. The presence of mummified dinosaur corpses, as well as "growth rings" in the teeth of some endothermic vertebrates¹⁸, provides evidence for seasonally dry conditions. By contrast the palaeosols of the Tullock formation have a weak microstructure, and tend to be poorly developed with thick organic layers and abundant siderite nodules, suggestive of a cooler climate and poor drainage. In combination with pollen data these palaeosols are indicative of a temperate flora dominated by conifers and situated in a swamp environment. A modern analogue would be bald cypress swamps that presently exist over portions of southeastern North America. Coals are thicker and better developed in the Tullock formations, consistent with the hypothesized existence of poorly drained, widespread wetlands.

The change in palaeosols is abrupt and occurs directly above the pollen break at the K/T boundary, although it is difficult to say how sudden the palaeosol change was because of a complex erosional hiatus at the boundary. The palaeosol change is generally consistent with an interpretation made by Smit and van der Kaars¹⁹ for palaeochannels in the Bug Creek area.

Evidence from both pollen and palaeosols suggest a change across the boundary to unstable conditions in the early Palaeocene, shown by channel downcutting and followed by the establishment of a new geomorphic equilibrium. Long-term palaeosol and pollen changes are compatible with base-level adjustments, possibly related to a drop, then a rise in sea level. The synchronous occurrence of an Ir anomaly, spheroids, pollen breaks and shocked quartz at the boundary is suggestive of an additional catastrophic destabilization.

The presence of abundant palaeosols places constraints on the degree of resorting of fossils within the Hell Creek Formation. For example, it has been suggested²⁰ that nearly all fossil material contained in the Hell Creek faunal-facies was reworked and transported. This is an overstatement, as palaeosols would be destroyed in frequently disturbed environments, and the widespread occurrence of palaeosols in association with large fossils (less commonly with microfossils) suggests that much, if not most of this facies was formed in place. The truncation and strong development of palaeosols at the K/T boundary, together with the lack of palaeosols associated with fossils of the Bug Creek faunal-facies is consistent with the suggestion¹⁹ that some of these fossils were transported.

A more detailed picture of the Hell Creek-Tullock palaeoecological transition awaits further analysis of data. We believe, however, that at the moment the asteroid impact hypothesis provides the most parsimonious explanation for changes in plant communities across the K/T boundary.

GUY D. LEAHY

2405 Bailey Hill Road,
Eugene, Oregon 97405, USA

MICHAEL D. SPOON

GREG J. RETALLACK

Department of Geology,
University of Oregon,
Eugene, Oregon 97403, USA

1. Bohor, B.F., Foord, E.E., Modreski, P.J. & Triplehorn, Science **224**, 867-869 (1984).
2. Alvarez, L.W., Alvarez, W., Asaro, F. & Michel, H.V. Science **208**, 1095-1108 (1980).
3. Smit, J. & Klaver, G. Nature **292**, 47-49 (1981).
4. Smit, J. & Kytte, F.T. Nature **310**, 403-405 (1984).
5. Bohor, B. Geology **12**, 695-696 (1984).
6. Alvarez, W. et al. Science **223**, 1135-1141 (1984).
7. Surlyk, F. & Johansen, M.B. Science **223**, 1174-1177 (1984).
8. Benson, R.H., Chapman, R.E. & Deck, L.T. Science **224**, 1334-1336 (1984).
9. Schimmelmann, A. & DeNiro, M.J. Earth, planet. Sci. Lett. **68**, 392-398 (1984).
10. Macellari, C.E. Geol. Soc. Am. Abstr. **16**, 582 (1984).
11. Hutton, C. Sixth Inter. Palyn. Conf. Abstr., 66 (1984).
12. Tschudy, R.H. et al. Science **225**, 1030-1031 (1984).
13. Pillmore, C.L., Tschudy, R.H., Orth, C.J., Gilmore, J.S. & Knight, J.D. Science **223**, 1180-1183 (1984).
14. Ganapathy, R. Science **220**, 1158-1161 (1983).
15. Findley, R. Nat. Geog. **159**, 3-65 (1981).
16. Napier, W.M. & Clube, S.V.M. Nature **282**, 455-459 (1979).
17. French, B.M. Science **226**, 353 (1984).
18. Bakker, R.T. in *A Cold Look at the Warm-Blooded Dinosaurs* (eds Thomas, R.D.K. & Olsen, E.C.), 351-462 (Westview, Boulder, Colorado, 1980).
19. Smit, J. & van der Kaars, S. Science **223**, 1177-1179 (1984).
20. Archibald, J.D. Geol. Soc. Am. Abstr. **16**, 432 (1984).

Fall out over atomic tests

John Simpson

Fields of Thunder: Testing Britain's Bomb. By D. Blakeway and S. Lloyd-Roberts. Allen & Unwin:1985. Pp.243. Hbk £10.95; pbk £3.95.

Clouds of Deceit. By J. Smith. Faber & Faber:1985. Pp.174. Hbk £8.95; pbk £4.95.

A History of British Atomic Tests in Australia. By J.L. Symonds. Australian Government Publishing Service:1985. Pp.593. A\$28.

Just Testing. By D. Robinson. Collins/Harvill:1985. Pp.595. pbk £5.95.

In January 1985 a unique event occurred. A foreign judicial commission came to Britain to cross-examine current and former scientific civil servants about events thirty years ago classified as military secrets, but now released. The investigations of this Australian Royal Commission into the short and long term consequences of British nuclear testing in Australia took place at a time of growing concern that participants in nuclear weapon tests in the atmosphere were more susceptible in later life to cancer, and certain other diseases, than the rest of the population. This concern was heightened by the attempts of those who had fallen ill to gain compensation for their suffering, disabilities and possible premature death from their parent governments; the refusal of certain of those governments to admit liability, and the organization of pressure groups to assemble evidence to demonstrate liability and to lobby for changes in compensation rules. It was against this background, including the publication next month in Australia of the Report of the Royal Commission, that these four books were written and can be evaluated.

Several obvious distinctions can be made between the books. Symonds was commissioned to write by the Australian Department of Resources and Energy as a person who held an official position in the Australian Atomic Energy Organization, and had privileged access to Australian and British documents whereas the other three are written by professional journalists and writers. It should therefore be no surprise that the theme of Symonds' work is that no unexpected events or irresponsible actions of states, scientists or officials occurred and, equally unsurprising, that the other three books appear to start from the assumption, fostered by long standing governmental secrecy, that there are human and state culpability and errors to be exposed when investigating British nuclear testing activities. Symonds' implicit conclusion is thus that there is little to be concerned about as a consequence of the tests, whereas the overt conclusion of the other three writers is the exact opposite.

There are other differences in the coverage. Smith and Symonds only cover events in Australia, while Blakeway/Lloyd-Roberts and Robinson deal with all

the British atmospheric nuclear tests, including those at Christmas Island. Symonds writes a simple chronology of events, using only government documents, while Blakeway/Lloyd-Roberts and Smith use historical sections, based upon Royal Commission documents and hearings, interviews with people involved in the tests and newspaper reports to argue a particular thesis.

Blakeway/Lloyd-Roberts argue that governments should now acknowledge that some, though not all, test veterans suffered as a consequence of the tests, whereas Smith presents a pot-pourri of accusations that the British media were subservient to the government of the day (an unsurprising view given her personal difficulties in having stories on nuclear testing published in the *Sunday Times*) and the need for nuclear disarmament and support for the test veterans against their parent governments. By contrast, half of Robinson's book consists of interview material gathered from servicemen involved in the tests. He attempts to view events from the perspective of the men rather than their commanders.

The literary style and presentation used in the volumes also differs appreciably. Symonds uses camera-ready copy, has appendices after most chapters in which important documents are reproduced and has a curious system of references to the side of the text. There are also elements of

repetition in the text. It is thus essentially an extended, detailed and exhaustive report, rather than a commercially produced book and its structure does not make for easy reading. Yet as a work of reference, it contains much more documented information than any of the other volumes.

Robinson, in half his book, uses an impressive technique of short quotations drawn from interviews with people who participated in the tests juxtaposed either on the basis of topics (for example eating facilities on Christmas Island) or experiences of specific tests. These quotations are interspersed with brief medical histories of each new individual introduced into the book. The technique gives a sense of what it was like to be there, of the ignorance of what was happening felt by those who were taking part and the way in which rumours about events spread.

The Blakeway/Lloyd-Roberts work is the more polished of the other two books, offering both greater detail and wider coverage than Smith. It is also the only one of the four with an index. It contains a number of glaring factual errors, however (e.g. on page 153 a bomb is dropped from an aircraft at 13,000 feet and explodes at 15,000 feet) and also is guilty of misrepresenting the proposals made in the mid-1950s to hold tests of initiator devices for nuclear weapons in Scotland. Thus on page 159 in a passage about the Christmas Island hydrogen weapon tests we are told that "the decision was bred from the same scientific confidence and political expediency that produced the idea that tests should take place over mainland Britain" yet the massive difference in the nature of the two types of tests is conveniently ignored.

In addition, when these authors find themselves confronted by lack of information, they occasionally resort to what can only be described as smear tactics. For example (page 177) "there is no indisputable evidence that large numbers of men

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Popperfoto

Britain's second atomic test at Woomera rocket base in Australia in January 1954 — watched here by Sir William Penny who led the team carrying out the tests.

at Christmas Island were unknowingly exposed to radiation" yet later in the same paragraph we are told "the hydrogen bomb tests were administered and carried out by the same authorities responsible for the Australian tests, and there is no reason to believe that the safety precautions changed for the better once the move had been made to the central Pacific". Similar remarks, however, could be made about all the books other than that of Symonds on the grounds that none of them has accurately reproduced the listing of UK atmospheric tests and their yields published by the Ministry of Defence in 1984.

Because three of the books draw on much the same body of information for their accounts of events, the main differences are in their perspective upon and interpretation of history, rather than any revelations about what happened. Yet a number of issues of substance do emerge. One is the crude techniques used for specifying the requirements for the initial test sites (the use of a map to see how far the nearest large town was from the US Nevada site and adding a safety margin). Another was the error of exploding the first device in the 1953 test series at a time when the meteorological conditions were unsuitable and the disputed consequences of this for both aborigines and the aircrew sent to tackle the resultant nuclear cloud.

Further issues include the degree to which procedures laid down for radiation safety were actually complied with; the extent to which aboriginal intrusions into the Maralinga test site went unreported in official documents; the degree to which unjustified risks of exposing personnel to radiation were taken in pushing the British nuclear weapon programme through to fruition in six years and finally, there was the impact of war-time agreements with Canada and the United States on the relationship between the United Kingdom and Australia.

One unresolved issue remains central to all these works: what assumptions can be made about the relationship between exposure to nuclear radiation, especially at low levels, and ill health and premature death. The official view remains that threshold values can be established to distinguish between those who are "safe" and those at risk. Opponents argue against this, largely on the grounds of statistical correlations. The neutral reader is likely to remain unconvinced either way on this issue by reading any of these books, but they will no doubt provide additional ammunition for a controversy with profound human, scientific and political dimensions which seems likely to continue indefinitely. □

John Simpson is Senior Lecturer in the Department of Politics, University of Southampton, Southampton SO9 5NH, UK. A second edition of his book, The Independent Nuclear State: The United States, Britain and the Military Atom, will be published next year by Macmillan, London.

The ups and downs of an antibiotic

Trevor I. Williams

Penicillin: Meeting the Challenge. By Gladys L. Hobby. Yale University Press: 1985. Pp. 319. \$30, £30.

THE timing of the events surrounding the appearance of penicillin is well-established. It was discovered by Alexander Fleming in 1928 and the crucial research demonstrating its remarkable therapeutic potential was carried out by Howard Florey, Ernst Chain, and a small group of colleagues at Oxford between 1939 and 1941, with large-scale production being established in the United States by 1944. In 1945 Fleming, Florey and Chain shared the Nobel Prize for Physiology or Medicine. Although there remained much to be learnt about penicillin, the peak effectively was passed forty years ago. It is surprising, therefore, that interest in the history of penicillin has lately intensified and still remains controversial. Since 1979 there have appeared an important new biography of Fleming; two official biographies of Florey; and an official biography of Chain is shortly to be published.

It might be supposed, therefore, that there would be little room for yet further work on the history of penicillin. This book by Gladys Hobby is different, however, in that it is a comprehensive and well documented review of penicillin as a drug and thus largely avoids the problems of personality that are likely to beset biographers.

The treatment is not wholly objective, however and, like others before her, Dr Hobby explores the problem of why Fleming not only failed to develop penicillin himself but quickly lost interest in it. She is surely right in supporting the view that a change in the climate of medical opinion was a significant factor. In 1928, there was much scepticism about the potential of chemotherapy, especially in St Mary's Hospital, which was strongly orientated towards vaccine therapy under Almroth Wright. By 1939, however, the demonstrable success of the sulphonamides had changed the attitudes of many people very considerably.

In commenting on Fleming's own plea that he failed through lack of technical support she makes one very curious statement. She attributes to Chain a comment "that without partition and paper chromatography, infrared and ultraviolet spectroscopy, nuclear magnetic resonance and electronic spin resonance — techniques not known in the 1920s and early 1930s — isolation of an unstable natural product such as penicillin would have been exceedingly difficult". It is hard to believe that Chain ever made such a statement,

for none of these techniques played any part in the original purification of penicillin for clinical trial. A possible explanation is that the author is confusing the isolation of penicillin with much later work on the chemical structure of penicillin to which, with an account of semi-synthetic penicillins, some 20 pages are devoted.

The first part of this book adds little to the description and interpretation of events in the UK save that it puts in better perspective the fundamental importance of N.G. Heatley's contribution to both the assay of penicillin and its extraction from crude culture medium. Its real strength lies in the second part, which deals with the development of penicillin in the United States after the visit by Florey and Heatley in the summer of 1941 to enlist the support of the American pharmaceutical industry. We are rightly reminded that even before that visit there had been some sporadic interest there in the phenomenon of antibiosis in general — Merck, for example, had supported S.A. Waksman at Rutgers University since 1939 — and penicillin in particular.

Dr Hobby herself, with M.H. Dawson and Karl Meyer at the College of Physicians and Surgeons in New York, experimented with crude penicillin very shortly after the publication of the first paper by the Oxford group in the *Lancet* of 24 August 1940. In 1944 she joined the pharmaceutical firm Pfizer as a microbiologist and is thus well qualified to recount the history of industrial production of penicillin in the USA. She also deals briefly with production in Canada, Holland, France, Germany and Japan. She brings out very clearly the dilemma of the American pharmaceutical industry during those critical years: heavy investment in fermentation plant might be rendered worthless almost overnight if the chemists came up with a viable synthesis — which in the event they did not.

One controversial issue is evaded, namely that process patents taken out by the US industry led to allegations in the UK Parliament and elsewhere that penicillin had been 'stolen' from Britain. Although these allegations were effectively disproved in a detailed inquiry initiated by Vannevar Bush in 1952, echoes of them are still heard. The overwhelming contribution of the USA to industrial production has encouraged popular belief there that penicillin was for all practical purposes an American discovery. Dr Hobby is at pains to dispel the illusion: "... penicillin was a British discovery — it was discovered in England, first studied in England — and it probably was one of Britain's major wartime contributions to society". □

Dr Trevor I. Williams, 20 Blenheim Drive, Oxford OX2 8DG, is editor of Endeavour. He worked in the Sir William Dunn School of Pathology, Oxford, 1942–5 and his biography 'Florey: Penicillin and After' was published by Oxford University Press in 1984.

Qwiffs and quantum leaps

Brian D. Josephson

Mind and the New Physics*. By Fred Alan Wolf. Heinemann: 1985. Pp. 342. £14.95.

IN THIS book the author expounds his ideas relating to a postulated connection between quantum mechanics and mind, a connection which is now beginning to be recognized as being possibly important by a number of scientists. Wolf starts from the conventional point of view that the quantum-mechanical wave function (or "qwiff" in his terminology) represents the potential for anything physical to become manifest, and regards the process of making decisions as one of interaction with the (potential) future. He postulates that we have control over the future by deciding what to observe, and thus determining which way the wave function should collapse. We capture possibilities by observing them and collapsing the wave function appropriately in the same way that a fisherman captures fish in his net. The author additionally discusses in a rather metaphorical way some information-processing aspects of quantum theory, as well as a large number of other, on the whole rather more questionable, ideas.

While the putative mechanism for being able to control the future has a certain degree of plausibility, conventional quantum theory would argue that the chance of *not* finding a particular possibility of interest should be taken into account as well as the one of actually finding it, these two possible outcomes being associated with wave function collapses in different directions. This would seem to lead to theoretical predictions distinctly different from those proposed by Wolf. His process for modifying the outcome would work only if a differential collapse mechanism were in operation, and would make the theory very close to E.H. Walker's theory of the control of physical reality by the will which was published in Vol. 3 of *Psychoenergetic Systems* (1979).

Unfortunately, the kind of attention to detail and use of logical methods of argument that might have made the author aware of matters such as the discrepancy between the mechanism he proposes and conventional theory is not a feature of this book. Its basic structure is one of informal expositions of standard theory, followed up by intuitions about their meaning in the author's scheme of things. Despite my sympathy with what Wolf is trying to do, it seems to me that the connections proposed, while occasionally being significant, have been arrived at mainly by a process of coupling together any two en-

titles that have some nebulous resemblance to each other (a typical example is the connection between Bose-Einstein statistics and love, discussed on pp. 152-154).

One can have little more confidence in the mathematical parts of the book, the illustrations for which have been produced using a computer apparently lacking graphics facilities, which hardly makes them easier to follow. Starting with an elementary howler in the example given to illustrate the addition of two wave functions on p. 191, which might be excused as a momentary lapse of attention, the author goes on to describe the equation $\langle 1/1 \rangle = 1$ as representing the law "The Mind is One". He then proceeds further to equations supposedly representing the characteristics of Enterprise, God and Satan, at which point the book starts to bear a distinct resemblance to some of the more dubious offerings from the public that arrive occasionally in one's mail box. It could be that there is some sense behind these strings of equations, but unfortunately Wolf does not consider it important to go into the details so that the critical reader can judge for himself.

The side effect of books such as this may be to cause other books and journal articles, which treat this subject in a more serious manner, to miss the attention they deserve. There come to mind Rupert Sheldrake's *A New Science of Life*, which has been criticized mainly on account of its unorthodoxy and not because of any errors or inconsistencies in its arguments; Fritjof Capra's *Tao of Physics*; D. Bohm and B.J. Hiley's article (*Fdn Phys. USA* 14, 255-274; 1984) which indicated that mind-like characteristics of nature may be inherent in the equations of quantum physics; and the two papers of H.P. Stapp (*Fdn Phys. USA* 12, 363-399; 1982 and 15, 35-47; 1985) on the connection between mind, matter and quantum mechanics. On the assumption that important new ideas may in fact be buried in Wolf's book, one must hope that a subsequent, more rigorous account will be written, directed at the scientist rather than at the non-scientist who seems to be the target of this one. □

Brian D. Josephson is Professor of Physics in the Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK.

Taking an aim

Gregory Gregoriadis

Red Blood Cells as Carriers for Drugs. Edited by J.R. DeLoach and U. Sprandel. Karger: 1985. Pp. 162. DM 125, \$44.25.

THE development of carriers to obtain optimal release of drugs where they are needed, and protection from hostile environments or premature loss, is a subject of growing interest and a hopeful bet in biotechnology ventures. Currently popular carriers include antibodies, polymers, liposomes and other colloids. With only a minority of workers favouring erythrocytes, what then is so special about them to warrant a meeting and a book? Several things according to the editors and the contributors. Most medical researchers are familiar with erythrocytes, a known quantity to be trusted, by clinicians especially (the danger of unwanted contaminants in other people's blood, is avoided by using the patient's own erythrocytes).

Erythrocytes can, in theory at least, optimize drug action in more than one way. For instance, they end up sooner or later in the reticuloendothelial system which is associated with a large number of microbial and other diseases, and could therefore carry drugs there. This aspect is discussed in chapters on the use of cells loaded with metal chelators which facilitate removal of stored iron, or with hydrolytic enzymes which can break down stored substrates in the tissues of patients with storage diseases. Further, (human) erythrocytes have a life span of over 100 days, an important feature for slow re-

lease of drugs into the circulation.

The book describes experiences with the release of antimicrobial and anticancer drugs from circulating erythrocytes. Inactivation of toxic metabolites entering erythrocytes by enzymes which have been introduced into the cells is discussed in chapters on asparaginase, rhodanase (co-trapped with sodium thiosulphate for cyanide detoxification), or urate oxidase-containing erythrocytes. Other applications include erythrocyte-mediated introduction of DNA and antibodies into eukaryotic cells and the enrichment of erythrocytes with allosteric effectors of haemoglobin or with 6-phosphate dehydrogenase.

Is this all too good to be true? Unfortunately yes. The editors and some of the authors recognize the formidable problems with the proposed uses, especially in terms of wide acceptability. For instance, there are already a number of simple and safe man-made carriers which can deliver drugs to the reticuloendothelial system efficiently. Further, release of drugs from circulating cells at a constant rate will probably be impossible to achieve. Perhaps the greatest advantage of erythrocytes over other systems is their long circulation time and their use in detoxification looks realistic enough. The book, discussing most aspects of the erythrocyte carrier and related technology in 19 short chapters, is the first one devoted entirely to this system and should prove useful to those interested in drug delivery. □

Gregory Gregoriadis is Head, Medical Research Council Group, Academic Department of Medicine, Royal Free Hospital School of Medicine, London NW3 2QG, UK.

* In the United States *Star Wave* published by Macmillan, New York: Hbk \$19.95, pbk \$7.95.

Putting flesh to the phantoms

Graham Burton & D.D.M. Wayner

Free Radicals in Biology and Medicine. By Barry Halliwell and John M.C. Gutteridge. Clarendon: 1985. Pp.346. £30, \$45.

THE transitory nature of free radicals, conferred on them by an unpaired valence electron, has made direct proof of their involvement in a chemical reaction a challenging task. This is especially so within the complex web of chemistry associated with biological systems. Nevertheless, there is now strong, albeit indirect, evidence that free radicals are involved in at least some normal and abnormal cellular processes, including, for example, photosynthesis, inflammation, cancer and toxic reactions to some drugs and pesticides. Research interest in free radicals in biology and medicine has now grown to the extent that there are entire conferences and journals devoted to the subject.

Anyone wishing to gain an appreciation of the state of knowledge in this diverse, interdisciplinary field, together with a feel for the difficulties involved in pinpointing the *in vivo* involvement of radicals, will welcome the arrival of this very readable book. Furthermore, the authors have realized that further progress will require

a better understanding of the chemistry of free radicals. Accordingly, they have made accommodations in the text and included an appendix on bonding and atomic structure to help those readers who lack the necessary background. However, the chemically naive may find the number of chemical reactions and structures daunting (and will not be helped by errors in the printed structures of some chemical intermediates).

More importantly, the speculative nature of some of the reaction schemes is not always indicated. For example, two mechanisms are presented for the trapping of the superoxide ion by the antioxidant vitamin E (p.62 and p.173); not only are they contradictory, but both seem unlikely because they involve electron and proton transfers which have never been demonstrated experimentally and are improbable on thermodynamic and kinetic grounds. A more likely possibility is that the small amount of the more reactive hydroperoxyl radical that is in equilibrium with superoxide ion at physiological pH will rapidly abstract a hydrogen atom from vitamin E, as is found to occur with other peroxyl radicals.

Reflecting the thrust of much of the research in this field, the authors devote considerable space to the potential roles played by the superoxide ion, hydrogen peroxide, the hydroxyl radical, iron and other transition metals, and the enzymes superoxide dismutase, catalase and glutathione peroxidase. They also describe the complex interplay of enzyme systems and the importance of maintaining sufficient concentrations of cofactors such as reduced glutathione and NADH. The reader, however, is left with the impression that the very reactive hydroxyl radical is the primary agent of destruction; this may be true, but it is also important to realize that the chain-carrying peroxyl radicals (which will almost certainly form as a result of hydroxyl radical attack on neighbouring molecules) can lead to a much larger increase in the initial damage. In this case, other antioxidant systems (for example vitamin E, vitamin C) become important.

There is no direct referencing within the body of the text (except where individual researchers are named) but each chapter is supplemented with an alphabetical list of selected references (up to and including 1984) for further reading. This system has its advantages, but will frustrate those pursuing specific points in the text.

Although the book is aimed at biologists and clinicians, it is a good starting point for anyone wishing to get a clear introduction to current aspects of free radical research in biology and medicine. It should, therefore, be of wide appeal. □

Graham Burton is an Associate Research Officer and D.D.M. Wayner a Research Associate in the Chemistry Division, National Research Council of Canada, 100 Sussex Drive, Ottawa, Ontario K1A 0R6, Canada.

Adding up parts of the past

I Grattan-Guinness

A History of Algebra from al-Khwarizmi to Emmy Noether. By B.L. van der Waerden. Springer 1985; Pp.271. DM 98, \$34.50.

ALGEBRA has long been a major component of mathematics, and so has attracted a good deal of interest among historians. This book occupies a special place, for its author has been one of the most distinguished algebraists of our century. After a chapter on medieval Muslim contributions, he takes the story through Renaissance Italy to Descartes, and then on to the algebraic structures that become a 19th-century concern: the work of Gauss and Galois, then the development of group theory, and the various algebraic systems which culminated in a sort of synthesis with van der Waerden's teacher, Noether.

As algebra has already been studied by many other historians, several parts of this book cover familiar ground: its novelties are drawn from recent research by others on various historical figures, which is used heavily and with full acknowledgement. While no major recasting of the history is thereby caused, the new details are very welcome. However, when one considers the interaction between algebra and other branches of mathematics, gaps in the story do emerge, which neither this nor other histories properly fills. I shall mention two.

One of the chief links for algebra has been with calculus, and its uses in applied mathematics. Indeed, a major issue there at times from the mid 18th century onwards was the possibility of securing some of these topics entirely within certain algebraic theories. While these attempts were successful, some important consequences followed for algebra: for example, they led to the early stages of functional equations and differential operators, the first theories in algebra explicitly divorced from interpretations in terms of length and number.

Partly as a result of these attempts, the 19th century saw the introduction of a variety of algebras in which laws other than the usual ones, suggested by arithmetic, were satisfied. For some reason the English became especially attracted to such studies, which gained collective prominence in their time, both in England and to some extent abroad. It is therefore rather astonishing that van der Waerden deals only with Hamilton, and that his book is completely silent on Babbage, Boole, Gregory, Herschel, De Morgan and Sylvester. The history of the history of algebra still has much life left in it. □

I. Grattan-Guinness is Reader in Mathematics, Middlesex Polytechnic, Enfield, Middlesex EN3 4SF, UK.

The life and work of J.B.S.

HALDANE

with special reference to India



vicki

Krishna R Dronamraju

This compelling biography of a courageous and complex scientist studies his contributions to the disciplines of genetics, evolution, biochemistry, physiology and statistics.

228 pp £14.90

ABERDEEN UNIVERSITY PRESS

Vaccination and herd immunity to infectious diseases

Roy M. Anderson & Robert M. May*

Department of Pure and Applied Biology, Imperial College, London University, London SW7 2BB, UK

* Biology Department, Princeton University, Princeton, New Jersey 08540, USA

An understanding of the relationship between the transmission dynamics of infectious agents and herd immunity provides a template for the design of effective control programmes based on mass immunization. Mathematical models of the spread and persistence of infection provide important insights into the problem of how best to protect the community against disease.

MUCH attention is now focused on the development and community-wide use of vaccines for the control of infectious diseases. This interest is a consequence of many factors; these include the worldwide eradication of smallpox by vaccination in the late 1970s¹, the current success of child immunization programmes for the control of measles, mumps, pertussis and rubella in the United States^{2,3}, the failure of frequent appeals by public health authorities in the United Kingdom to raise vaccination rates that are at present low by developed-country standards⁴⁻⁷, the Expanded Programme on Immunization (EPI) of the World Health Organisation (WHO) for the control of a variety of serious childhood viral and bacterial diseases in the developing world⁸⁻¹², and the rapid advances in molecular biology and biotechnology which promise new vaccines for the future¹³⁻¹⁶. The main thrust of research today is at the molecular level, where it is stimulated by the urgent need to develop vaccines for the major killing diseases of the Third World such as malaria¹⁷⁻¹⁹, and to combat new infections such as the current epidemic of AIDS (acquired immune deficiency syndrome), caused by the recently isolated human T-lymphotropic virus (HTLV-III)^{20,21}.

The development of a safe, effective and cheap vaccine, however, is only a first step (albeit an essential one) towards community-wide control. Epidemiological, economic and motivational issues are at least as important as technological ones. A vaccine for measles (a very cost-effective immunization¹²) has been available since the late 1960s, yet the infection remains one of the world's major causes of child mortality^{8,22}. Concomitant with research on vaccine development, there is a need for an improved understanding of how best to use vaccines to protect the community as well as the individual. The persistence of infectious disease within a population requires the density of susceptible individuals to exceed a critical value such that, on average, each primary case of infection generates at least one secondary case. It is therefore not necessary to vaccinate everyone within a community to eliminate infection; the level of herd immunity must simply be sufficient to reduce the susceptible fraction below the critical point. The central epidemiological questions are thus: what proportion of the population should be vaccinated to achieve elimination (in a local programme), eradication (in a global programme) or a defined level of control? How is this affected by demographic factors (for example, birth rate)? What is the best age at which to immunize? How does mass immunization affect the age distribution of susceptible individuals, particularly in those classes most at risk from serious disease, and how important is genetic and spatial heterogeneity in susceptibility to infection (or response to immunization) to the creation of effective herd immunity²³⁻²⁶?

To answer these questions, we need to understand the interaction between the transmission dynamics of the disease agent and the level of naturally acquired (or artificially created) immunity to infection. This relationship is complex and depends on such factors as the precise course of infection within an individual, the demography of the host population, the duration

of acquired immunity and maternally derived protection, age-related changes in the degree and intimacy of contacts among people, and the prevailing levels of genetic, spatial and behavioural heterogeneity in susceptibility/resistance to infection. Intuition built on many years of practical epidemiological experience often fails to predict the outcome of a particular vaccination programme. Mathematical models which clearly and accurately define the details of the association between disease agent and host (at the individual and population levels) can help to overcome such difficulties. This article reviews recent theoretical and empirical work on the transmission dynamics of infectious disease. Emphasis is placed on the relevance of such research to the design of disease control programmes based on vaccination.

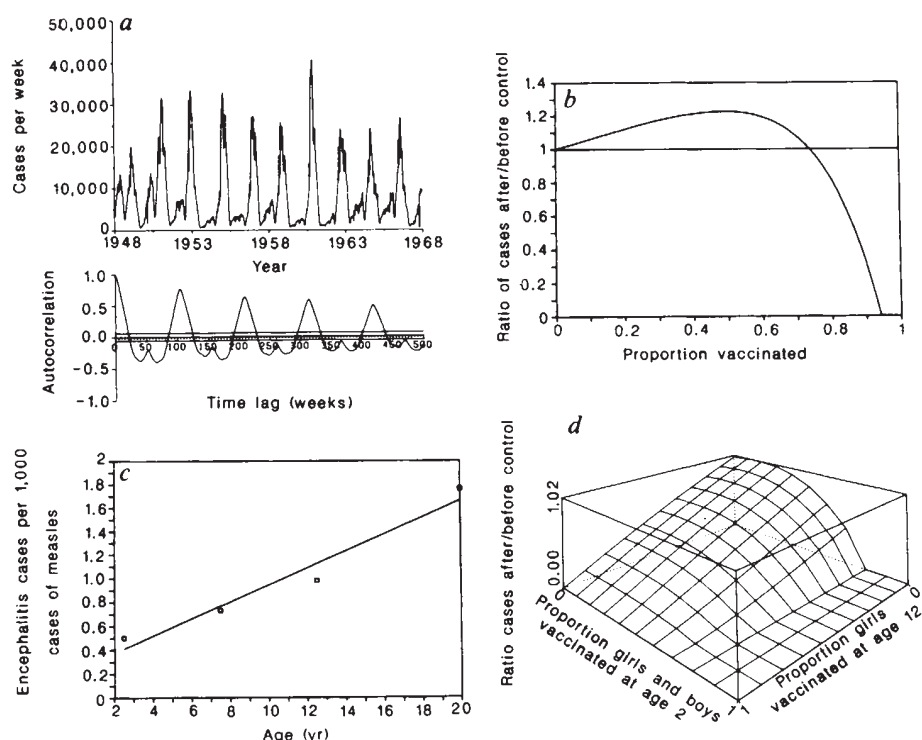
General theory

Models for the transmission dynamics of directly transmitted viral and bacterial infections are typically compartmental in structure²³⁻³⁵. They represent changes, with respect to age, a , and time, t , in the populations of infants protected by maternally derived antibodies ($M(a, t)$), susceptible individuals ($X(a, t)$), infected individuals who are not yet infectious ($H(a, t)$), infectious people ($Y(a, t)$), and people who have recovered and are immune from subsequent infection ($Z(a, t)$). The total population is $N(a, t) = M(a, t) + X(a, t) + H(a, t) + Y(a, t) + Z(a, t)$. The deterministic equations describing the flows from one class to another are of the general form^{24,25,27}

$$\begin{aligned} \frac{\partial M(a, t)}{\partial t} + \frac{\partial M(a, t)}{\partial a} &= -[\mu(a) + d]M(a, t) \\ \frac{\partial X(a, t)}{\partial t} + \frac{\partial X(a, t)}{\partial a} &= dM(a, t) \\ &\quad - [v(a, t) + \lambda(a, t) + \mu(a)]X(a, t) \\ \frac{\partial H(a, t)}{\partial t} + \frac{\partial H(a, t)}{\partial a} &= \lambda(a, t)X(a, t) \\ &\quad - [\mu(a) + \sigma]H(a, t) \\ \frac{\partial Y(a, t)}{\partial t} + \frac{\partial Y(a, t)}{\partial a} &= \sigma H(a, t) - [\mu(a) + \gamma]Y(a, t) \\ \frac{\partial Z(a, t)}{\partial t} + \frac{\partial Z(a, t)}{\partial a} &= \gamma Y(a, t) + v(a, t)X(a, t) \\ &\quad - \mu(a)Z(a, t) \\ \frac{\partial N(a, t)}{\partial t} + \frac{\partial N(a, t)}{\partial a} &= -\mu(a)N(a, t) \end{aligned} \quad (1)$$

Here $\mu(a)$ denotes age-related human mortality, $1/d$ is the average duration of protection provided by maternal antibodies, $1/\sigma$ and $1/\gamma$ are the average incubation and infectious periods respectively, and $v(a, t)$ records the age- and time-specific vaccination schedule. Mortality explicitly associated with the

Fig. 1 *a*, Time series analysis of the case reports for measles in England and Wales over the pre-vaccination period 1948–68. The top graph denotes the recorded cases and the bottom graph the serial correlogram⁵¹. Note the significant yearly (seasonal) and 2-yearly peaks in incidence (the horizontal axis in the bottom graph denotes weeks). *b*, Model predictions of the equilibrium (the state to which the system settles in the long term) ratio of the number of cases of mumps in the age range 14–50 yr (the age range in which males are at risk to serious disease resulting from infection) under the impact of a vaccination programme in which *p* of each yearly cohort are vaccinated at age 2 yr, divided by the number of cases in the same age range before vaccination. If the ratio is above unity, the control programme has a detrimental effect and vice versa. The average age at infection before control was assumed to be 6.5 yr (appropriate for the United Kingdom) and the calculations were based on a model defined in ref. 24. The straight line denotes the ratio of unity. *c*, Age-specific incidence of reported measles encephalitis cases (defined per 1,000 cases of measles) in the United States in 1973–79¹⁶. *d*, Similar to *b* but representing the ratio of cases after vaccination divided by cases before vaccination for rubella in the United Kingdom in the age range 16–40 yr. The graph denotes model predictions (equilibrium values) for a two-stage vaccination policy involving vaccination of only girls at age 12 yr (the UK policy) and vaccination of boys and girls at age 2 yr (the US policy)^{25,67}. Note that the addition of a second-stage US policy over the current UK policy is only of substantial benefit if moderate to high levels of vaccination coverage are achieved at 2 yr of age. The average age at infection before control was assumed to be 9–10 yr of age^{25,67}.



infection has been ignored. The term $\lambda(a, t)$ denotes the 'force' or rate of infection, which under the 'mass action' assumption adopts the form^{24–30} $\lambda(a, t) = \int_0^\infty \beta(a, a') Y(a', t) da'$. Here $\beta(a, a')$ represents the transmission coefficient for infected individuals in age class a' who come into contact with susceptible individuals in age class a (refs 25, 30). The boundary conditions of equations (1) depend on the demography of the community. The structure of the model may be modified to mimic the transmission of a wide variety of infections^{31–34}, including those with complex life cycles (for example, malaria or yellow fever^{35–37}) and those involving carriers (such as tuberculosis and hepatitis B³⁸) or macroparasites (such as helminths) where note must be taken of the distribution of parasite numbers within the human community^{39,40}.

Insights into applied problems of disease control or into general ecological and epidemiological questions can be obtained through analytical or numerical studies of equations (1).

Cycles in incidence

For most common viral and bacterial infections of childhood, the model predicts large-amplitude, slowly damping oscillations in disease incidence^{23–30}. Several factors can tend to perpetuate these oscillations indefinitely; these are stochastic effects^{41–44}, age-dependent transmission rates²⁵, incubation periods of a discrete nature^{45,46} and seasonality in contact^{27,47–50}. The average period of oscillations, T , is approximately given as²³: $T \approx 2\pi(AK)^{1/2}$, where A is the average age at infection and K is the generation time of the infection ($K = 1/\sigma + 1/\gamma$).

The observed periods of such cycles (which agree well with prediction) can be determined accurately by time-series techniques (Table 1a, Fig. 1a⁵¹); the longer-term cycles are usually compounded with short-term seasonal effects.

Vaccination and herd immunity

The basic reproductive rate or transmission potential of an infection, R_0 , is the average number of secondary cases produced by one primary case in a wholly susceptible population^{33,34}. Clearly, an infection cannot be maintained unless R_0 exceeds

unity. Within communities in which the infection is endemic, R_0 is inversely correlated with A , the average age at infection^{23–37} ($R_0 \approx B/(A - D)$). This relation holds both for developed countries (growth rates essentially zero) and for developing countries (with their growing populations)⁵². Here B is the reciprocal of the finite birth rate (usually expressed as births (yr^{-1}) per 1,000 of population), and D is the duration of maternally derived immunity. The value of R_0 within a community is critically dependent on social and behavioural factors (the contact component of β which influences λ and hence A), on the biology of the disease agent and host (the infectiousness and susceptibility components of β) and on the demography of the community (value of B)²⁵. The marked difference between developed and developing countries in the average ages at which children typically acquire infections such as measles in urban communities is a consequence of variations in social, behavioural and demographic factors (Table 1b).

To eradicate an infection by mass immunization it is necessary to reduce the value of R_0 below unity^{23–25,53}. This can be achieved by vaccinating a proportion, p , of each cohort of the population at age V (where $V > D$; vaccination during the period of maternally derived protection usually fails to protect the child adequately from subsequent infection^{54,55}) provided^{23,25} $p > [1 - A/B]/[1 - V/B]$. The infection can never be eradicated if $V > A$ (that is, $p > 1$). The value of p is minimized by vaccinating at as young an age as possible, given the constraint of $V > D$. These conclusions are derived under the assumption that the rate of transmission within the community is fairly homogeneous. In practice, this assumption is often violated^{25,30}, but simple theory provides a useful guide, usually setting an upper limit to the levels of vaccination required to eliminate many common childhood infections²⁵. It forms the basis of estimates suggesting that 92–96% of children must be vaccinated to eliminate measles and pertussis, 84–88% to eliminate rubella and 88–92% to eliminate mumps in Western Europe and the United States; these estimates assume that vaccination takes place just after the wane of maternal antibody protection^{23–25}. Recent experience in the United States suggests that high levels of cohort immunization are required (as indicated by theory)

Table 1 Infections reported throughout the world

a			b		
Infectious disease	Inter-epidemic period (yr)	Geographical location and time period	Infectious disease	Average age A (yr)	Geographical location and time period
Measles	2	England & Wales, 1948-68	Measles	5-6	USA, 1955-58
	2	Aberdeen, Scotland, 1883-02		4-5	England & Wales, 1948-68
	2	Baltimore, USA, 1900-27		2-3	Morocco, 1962
	2	Paris, France, 1880-10		2-3	Ghana, 1960-68
	1	Yaounde, Cameroun, 1968-75		2-3	Pondicherry, India, 1978
	1	Ilesha, Nigeria, 1958-61		1-2	Senegal, 1964
Rubella	3-5	Manchester, UK, 1916-83	Rubella	1-2	Bangkok, Thailand, 1967
	3-5	Glasgow, Scotland, 1929-64		9-10	Sweden, 1965
Parvovirus (HPV)	3-5	England & Wales, 1960-80		9-10	USA, 1966-68
Mumps	3	England & Wales, 1948-82		9-10	Manchester, UK, 1970-82
	2-4	Baltimore, USA, 1928-73		6-7	Poland, 1970-77
Poliomyelitis	3-5	England & Wales, 1948-65		2-3	Gambia, 1976
Echovirus (type II)	5	England & Wales, 1965-82	Chickenpox	6-8	USA, 1912-28
Smallpox	5	India, 1868-48	Poliovirus	12-17	USA, 1955
Chickenpox	2-4	New York City, USA, 1928-72	Pertussis	4-5	England & Wales, 1948-68
	2-4	Glasgow, Scotland, 1929-72		4-5	USA, 1920-60
Coxsackie virus (type B2)	2-3	England & Wales, 1967-82	Mumps	6-7	England & Wales, 1975-77
Scarlet fever	3-6	England & Wales, 1897-78		6-7	Netherlands, 1977-79
Diphtheria	4-6	England & Wales, 1897-79			
Pertussis	3-4	England & Wales, 1948-85			
<i>Mycoplasma pneumoniae</i>	4	England & Wales, 1970-82			

c	Age groups (yr)					Proportion to be immunized for eradication (p)
	Force of infection λ (yr ⁻¹)	0-5	5-10	10-15	15-20	20-75
Case 1	0.2	0.2	0.2	0.2	0.2	0.94
Case 2	0.18	0.56	0.2	0.1	0.1	0.89

a, The inter-epidemic period, T , for various viral and bacterial infections^{23-25,51,99,115}. b, Average age at infection (yr), A , for different diseases in different localities^{23,24,34,99,116}. c, The impact of age-related changes in the force of infection, λ (yr⁻¹), on the critical proportion of the population p , to be vaccinated to eradicate measles in England and Wales²³⁻²⁶ (case 1, constant; case 2 age dependent).

to substantially reduce the incidences of common childhood infections.

The vaccination of cohorts must be continued for many decades before the full impact of a programme is manifest; long-term suppression or interruption of transmission requires continual cohort vaccination as long as the infection remains endemic in other communities or countries. The only infection to be eradicated worldwide is smallpox. This success (announced by WHO in 1977) generated some optimism concerning the feasibility of eradicating other major infections, such as measles and pertussis^{8,9,56-58}. Such optimism, however, may be misplaced, given the low communicability of smallpox (low R_0 and high average age at infection, A , before widespread immunization) in comparison with infections such as measles, the ease with which smallpox infection can be diagnosed, the practical simplicity of inoculation procedures and the stability of the smallpox vaccine under primitive storage conditions^{1,59,60}.

Successful immunization programmes may themselves create health problems, since vaccination almost invariably carries some risk (usually extremely small) to the individual. At the start of a mass-immunization programme, the probability of serious disease arising from vaccination is usually orders of magnitude smaller than the risk of serious disease arising from natural infection. As the point of eradication is approached, the relative magnitude of these two probabilities must inevitably be reversed. The optimum strategy for the individual (not to be vaccinated) therefore becomes at odds with the needs of society (to maintain herd immunity). This issue—which was central to the decline in uptake of whooping cough vaccine in the United Kingdom during the mid- to late-1970s—can be overcome by legislation to enforce vaccination (as in the United States), but its final resolution is only achieved by global eradication of the disease agent (so that routine vaccination can cease)⁶¹.

Mass immunization at levels below that required for elimination acts to reduce the number of cases of infection (and hence the force of infection, λ). However, both theory²³⁻²⁶ and observation^{62,63} suggest that it has little, if any, impact on the total number of susceptible individuals. If the infection is endemic, its effective reproductive rate is unity—each case produces, on average, one other case—and the fraction susceptible remains roughly constant ($\sim 1/R_0$), independently of whether suscepti-

bility is lost by vaccination or by natural infection. Reduction of transmission via vaccination, however, usually increases the inter-epidemic cycle period and the average age at infection^{23-26,51}. This latter manifestation can be worrying if the risk of serious disease resulting from infection increases with age (Fig. 1c). Whether this issue is important depends on the quantitative details of such factors as how risk changes with age, the average age at which the vaccine is administered, and how the force of infection changes with age before widespread immunization^{25,26} (Fig. 3a). Rubella and mumps are of particular concern because of the risk of congenital rubella syndrome (CRS) in infants born to mothers who contracted rubella in their first trimester of pregnancy^{64,65} and the occurrence of orchitis and the associated risk of sterility in post-pubertal males following mumps infection⁶⁶ (Fig. 1b, c). The problems surrounding mass vaccination against rubella (to control CRS) have resulted in the adoption of different vaccination programmes in different countries. In the United Kingdom, girls only are vaccinated at an average age of around 12 yr, so as to allow rubella virus to circulate and create naturally acquired immunity in the early years. By contrast, in the United States both boys and girls are vaccinated at around 2 yr of age, with the aim of eliminating rubella. Theory suggests that the US policy is best if very high levels of vaccination (80-85% of each yearly cohort) can be achieved at a young age, while the UK policy is better if this cannot be guaranteed^{24,67}. A mixed US and UK policy is predicted to be of additional benefit over the UK single-stage policy if moderate to high levels of vaccine uptake among boys and girls can be achieved at a young age ($>60\%$)⁶⁷ (Fig. 1d).

Epidemiological data

The average age at infection, A , is central to most estimates of the intrinsic transmission potential of an infection, to the calculation of the desired level of vaccination coverage for elimination, and to theoretical assessment of how the risk of serious disease caused by infection may change with age under the impact of mass vaccination. For infections which are thought to induce life-long immunity against disease (for example, measles, mumps, rubella and pertussis), the value of A can be estimated (see Table 1b) either from age-stratified case notification records or, more accurately, from serological profiles (Fig. 2).

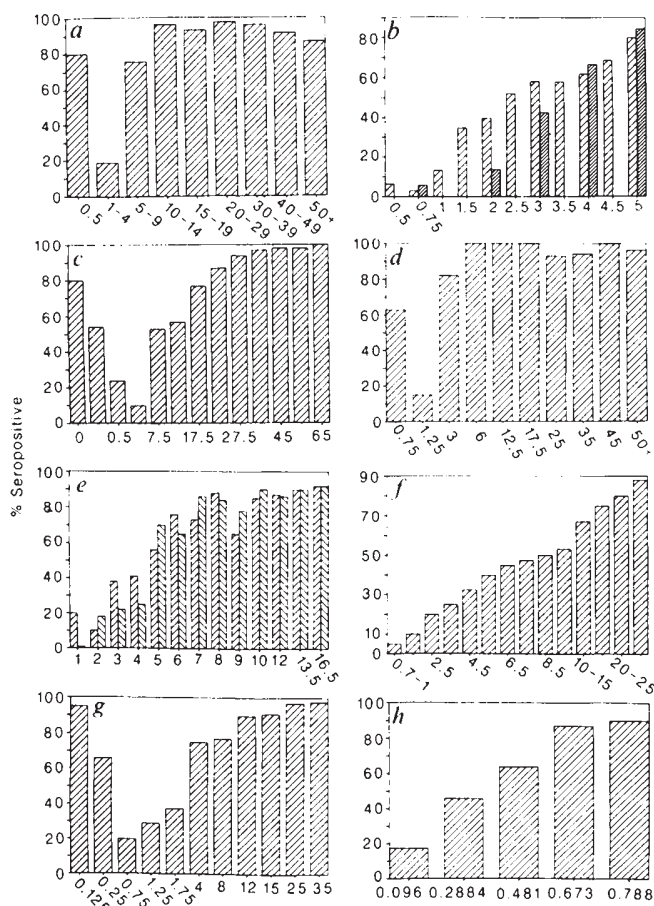


Fig. 2 Age-stratified (horizontal) serological profiles of the percentages of people in different age groups with antibodies to various infectious disease agents in developed and developing countries. *a*, Measles in the United States in 1955-58 (New Haven, Connecticut)¹¹⁷. *b*, Measles in India¹¹⁸ (left-hand histograms) and Nigeria¹¹⁹ (right-hand histograms). Note the rapidity with which the proportion seropositive rises with age when compared with *a*. *c*, Rubella in England and Wales²⁴ in the 1960s and 1970s. *d*, Rubella in the Gambia in the late 1970s¹²⁰. Again note the rapidity of the rise in seropositivity with age when compared with *c*. *e*, Mumps in England and Wales (left-hand histograms) and the Netherlands (right-hand histograms)^{121,122}. *f*, Polio in the United States in 1955³⁴. *g*, Hepatitis B (HBV) in Senegal¹²³. The vertical axis denotes the percentage with HBV markers. *h*, Malaria (*Plasmodium falciparum*) in Nigeria¹¹⁵. The vertical axis denotes the cumulative percentage of children who have experienced infection. Note that in *a-h* the scales of the x-axes are not always identical and in certain cases record unequal age class divisions (a nonlinear scale).

The interpretation of the data, however, is beset with many problems, such as age-specific biases in notification records^{68,69}, and a reduced ability to detect antibodies in adults who experienced infection in early childhood⁷⁰.

Heterogeneity in transmission

Conventional epidemiological theory is founded on the 'mass action' principle of transmission (see equations (1)), in which infection is assumed to spread via contacts within a homogeneously mixing community⁷¹. Recent research has begun to take account of the complications induced by heterogeneity in transmission arising from age-related, genetic, spatial or behavioural factors^{25,30,72-77}.

Age-dependent factors. Analyses of case notification records and serological profiles suggest that, for many common infections (measles, rubella and pertussis), the per capita rate of infection ($\lambda(a, t)$) depends on the ages of susceptible individuals, changing from a low level in the 0-5-yr age classes, via a high level in the 5-15-yr age classes, back to a low level in the adult age classes (Fig. 3a)²⁵. This is of interest both because it reflects behavioural attributes of human communities and because of

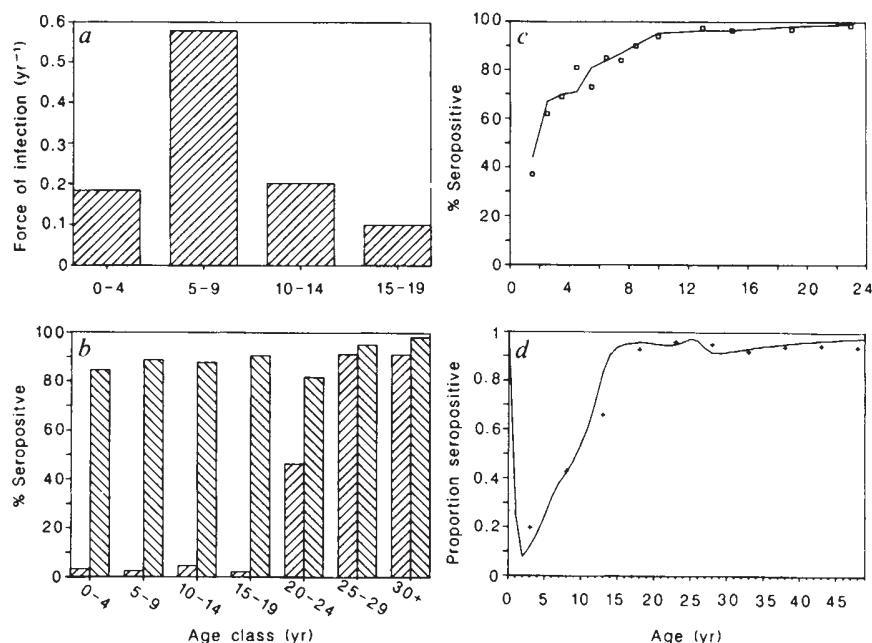
its impact on the level of vaccine-induced herd immunity required to eliminate an infection^{25,67,75}. The high values of λ observed in the 5-15-yr-old classes (Fig. 3a) are thought to arise as a consequence of frequent and intimate contacts within school environments.

Analyses of the significance of age-related changes in contact with infection are crude at present, due to the absence of detailed information on 'who acquires infection from whom'²⁵. Models which incorporate high degrees of contact within school-age classes (and low contact in the very young and in adult age groups), however, yield predictions of temporal changes in disease incidence and of age-related changes in serology in vaccinated versus unvaccinated communities which closely parallel observed trends²⁵ (Figs 3c, d, 4a-c). Recent numerical studies suggest that predictions based on the assumption of homogenous mixing tend to overestimate the critical level of vaccination coverage for elimination by roughly 5% for infections such as measles and pertussis in the United Kingdom and United States (Table 1c)^{25,30}. For these infections, analyses based on apparent age-related changes in contact with infection (Fig. 3a) yield estimates of approximately 90% coverage at 1-2 yr of age²⁵. These predictions, however, must be accepted with caution at present, because the observed low rates of infection in late teenage and adult age groups may be artefacts arising from biases in case reporting, from the insensitivity of serodiagnostic tests, or from genetic variability in susceptibility to infection⁷⁸. These caveats are underlined by the fact that no such marked age-related changes in the per capita rate of infection are found in epidemics within virgin populations on islands (largely susceptible to infection before the epidemic⁷⁹⁻⁸¹) (Fig. 3b). At present it seems prudent to accept the higher vaccine coverage levels estimated by homogeneous mixing models²⁵.

Genetic factors. Genetic factors are important in determining host susceptibility and resistance to infectious agents⁷⁸. Genetically based variability in the level and duration of antibody production following infection would seriously complicate the interpretation of age-serological profiles⁸². Associations between HLA type and antibody titres following vaccination against rubella and measles provide evidence for the existence of such complications^{83,84}. Little is understood, at present, about the importance of such factors in the design of vaccination programmes. In the future, however, community-based diagnostic programmes aimed at the identification of immunodeficiency could make it possible to target vaccination, for certain less-common but serious viral and bacterial diseases, to those children genetically predisposed to infection and severe morbidity⁸⁵.

Spatial factors. Inhomogeneity in the spatial distribution of hosts, with some people living in dense aggregate and others living in isolated or small groups, can lead to heterogeneity in transmission rates. This, in turn, can result in the transmission potential of an infection (R_0) being greater on average than suggested by estimation procedures which assume spatial homogeneity^{86,87}. Under these circumstances, the optimal solution appears to involve 'targeting' vaccination coverage in relation to group size, with dense groups receiving the highest rates of vaccination⁷⁴. The optimal programme is defined as that minimizing the total, community-wide number of immunizations needed for elimination or a defined level of control. This strategy reduces the overall proportion that must be vaccinated to achieve elimination, compared with that estimated on the assumption of spatial homogeneity. This conclusion has practical significance for the control of infections such as measles and pertussis in some developing countries, where rural-urban differences in population density tend to be much more marked than in developed countries. It is probable that in many regions of Africa and Asia, diseases such as measles cannot persist endemically in rural areas without frequent movement of people between rural and urban regions ($R_0 < 1$ in rural areas). Under these circumstances, disease control might be achieved in both types of region by high levels of mass immunization in the urban centres alone.

Fig. 3 *a*, Age-specific change in the rate, or force, of measles infection (λ) in England and Wales before mass vaccination²⁵. *b*, Percentages seropositive for rubella antibodies in various age classes on St Paul's Island, Alaska, before (left-hand histograms) and after (right-hand histograms) an epidemic of rubella in 1963⁷⁹⁻⁸¹. The population (~400 people) had not experienced rubella infection (before 1963) for 22 yr. Note that in the age range 0-19 yr, the rate of infection of susceptible individuals was similar in each age class. *c*, Comparison of model predictions (equations (1)) of the age-serological profile for measles antibodies in England and Wales with observed patterns. The observed data (squares) were collected in the period 1979-82. Model predictions were based on an average age of infection of 4.5 yr before control and the age-specific vaccination rates in the United Kingdom over the period 1968-82²⁵. *d*, Comparison of model predictions of the age-serological profile for rubella antibodies in females in England and Wales with the observed pattern (crosses) in 1984⁶⁷. Predictions were based on an average age of infection of 10 yr before control and the age-specific vaccination rates in the United Kingdom (vaccination of teenage girls between the ages of 10 and 15 yr) over the period 1970-84.



Behavioural factors. These are of great importance as determinants of disease transmission and provide a source of significant heterogeneity. The sexually transmitted infections provide a good example because the probability distribution of sexual promiscuity (defined as the number of different sexual partners per unit of time) is highly skewed. It seems very likely that the highly sexually active individuals ($R_0 \gg 1$) in the tail of the distribution are responsible for maintaining transmission within the community as a whole⁸⁸⁻⁹⁰. Drug treatment and surveillance focused on these 'superspreaders' is therefore highly beneficial, given that R_0 may be below unity for the remaining population. If vaccines are developed for HTLV-III, selective immunization of those most at risk⁹¹⁻⁹³ (haemophiliacs, intravenous drug abusers, persons of Central African or Haitian origin and homosexual males) and those predisposed to serious clinical infection, may be a practical option for community-wide control.

Behavioural factors are also important in helping to cause seasonal fluctuations in the incidence of many common viral and bacterial infections^{49,75,94}. The timings of school terms and vacation periods has a marked influence on temporal patterns of measles and pertussis cases in the United Kingdom and West Germany, for example.

Developing countries

Many infections that in developed countries are primarily a source of morbidity are major cause of mortality in the developing world (mainly as a consequence of malnutrition, although genetic components may also be involved)^{9,22,95}. Measles, for example, is estimated to cause more than one million child deaths annually^{9,22}. Although there is much current interest in worldwide immunization programmes for the control of such infections, the problems of creating sufficient levels of herd immunity for elimination should not be underestimated, given the logistic difficulty of vaccine storage and delivery in poor countries. Even if such practical issues are resolved by effective cold-chains, the development of heat-stable vaccines, and aerosol delivery techniques (which are more effective in the immunization of infants with maternal antibodies)⁹⁶⁻⁹⁸, the feasibility of blocking transmission by mass vaccination in areas where the average age of infection is often as low as 1-2 yr of age is questionable (Table 1b).

Consider a large urban centre in India or Africa, where the annual birth rate is around 40 per 1,000 population and the average age of measles infection is 2 yr (Fig. 2). Vaccination must take place after the maternally derived protection has

waned (6 months) and before the age of 2 yr to achieve elimination. Calculations suggest an R_0 value of around 17, which is similar to that for measles in England and Wales (with a birth rate of 12 per 1,000 and an average age of infection around 4.5-5 yr). Simple theory suggests that something like 96% of each yearly cohort of children would have to be effectively immunized by age 1 yr ($V=1$) to attain elimination. Coverage could be lower in rural areas but, as suggested in the section on spatial heterogeneity, an optimal control programme would have to focus on the urban centres. These estimates question the feasibility of measles elimination in many developing countries at this time and argue that, as a first priority, worldwide programmes should concentrate on controlling morbidity and mortality among those most at risk^{52,99}. Moderate to low levels of vaccination coverage, however, are beneficial; they raise the average age at infection and thereby lengthen the age 'window' in which vaccine can be administered⁹⁹, and reduce morbidity and mortality (young infants appear to be more at risk than older children).

New vaccines

The coming decade will see the development and use of a number of new vaccines, along with the expansion of community-based programmes of mass immunization using the currently available vaccines. The development of safe, effective and cheap vaccines is clearly the first priority, but, once these objectives are achieved, the problem of community-wide control requires careful attention. In particular, age-specific serological profiles of the community need to be obtained both before and during any programme of mass vaccination.

An example is provided by the recent proposal to initiate a mass-vaccination campaign for the control of hepatitis B virus (HBV) in Taiwan¹⁴ by a serum-derived vaccine (a genetically engineered vaccine is also available^{100,101}; Fig. 2). The epidemiology of HBV is made complicated by chronic infection in immunodeficient individuals who become carriers of the virus (with a high associated risk of inflammatory disease and malignancy)¹⁰², and who constitute an important reservoir of infection. The desirability of mass immunization is clear when viewed as a method for reducing the incidence of carriers (the risk of becoming a carrier appears to be much greater in children than in adults¹⁰³), but less clear given the observation that serious symptoms of disease are more likely to occur in adults than in children. Serological surveys in high-risk regions (South-East Asia and Africa) indicate that HBV typically has a low R_0 within the population as a whole (when compared with measles).

Moderate levels of vaccination coverage should therefore have a significant impact on disease incidence^{23,102,104} (Fig. 1d). The calculation of R_0 is made complicated by the presence of carriers ($R_0 = R_{01}f + R_{02}(1-f)$, where f is the fraction of carriers and $(1-f)$ the fraction of non-carriers) and a component of perinatal (=vertical) transmission⁴². The maintenance of transmission ($R_0 > 1$) is probably due to the carriers alone ($R_{01} > 1$, $R_{02} < 1$) given that they are often a significant fraction of the community (with f values ranging from 0.1% to 2% in the United States and most European countries, to as high as 50% in certain Pacific islands (determined in part by the ethnic composition of a community¹⁰²)).

Vaccines for viral infections such as cytomegalovirus, rhinovirus, adenovirus, rotavirus, respiratory syncytial virus and parainfluenza virus may provide sufficient protection to the individual to be considered for community-wide use in the future, provided their public health significance is of sufficient importance relative to other priorities in health care. The existence of many distinct antigenic types within certain groups of these viruses (for example, adenovirus and rhinovirus), however, present problems for vaccine development.

Recent work has raised the hope that vaccines may be developed for the major parasitic infections such as malaria (*Plasmodium falciparum*) and certain helminth species¹⁰⁵. The problems, however, are formidable because of the antigenic complexity of protozoa and helminths (in comparison with viruses and bacteria)¹⁰⁶, sequential changes in surface antigens during development within the host (for example, the filarial worms)¹⁰⁷, antigenic variation (whether due to the parasite's ability to express many variable antigen types, as in the case of the African trypanosomes^{108,109}, or as a consequence of population genetic (=strain) variability, as in the case of the malaria parasites¹¹⁰), and the failure of man naturally to develop fully protective immunity under conditions of repeated exposure to infection¹¹¹. Vaccines must be more effective in stimulating the host's immunological defences than the parasite itself if they are to overcome the evasion mechanisms that certain parasitic species have evolved¹¹². Combined with these problems is the suspicion that heterogeneity in immunological responsiveness is an important determinant of the observed aggregation in the distribution of parasite numbers per person (such that 15% of the hosts will often harbour 90% or more of the parasites)^{39,40,113,114}. Predisposition to heavy or repeated parasitic infection is often a consequence of behavioural or social factors but laboratory studies increasingly point to the importance of genetic and nutritional components¹¹². An effective vaccine must therefore be able to protect both immunocompetent and immunodeficient individuals, the latter being those most likely to suffer serious disease and to maintain parasite transmission.

The ideal malaria vaccine will probably contain antigens from different malaria species, strains and developmental stages (sporozoite, asexual and gamete) so as to reduce mortality and morbidity, and to block transmission^{18,19}. But once the technical problems in vaccine development have been overcome, the issues of community control must take priority. The creation of a level of herd immunity high enough to eliminate malaria appears impractical in endemic regions where the average age at first infection is typically around 3–6 months of age ($R_0 = 50$ –100, which implies 99% coverage of the community before the age of 3 months with a vaccine which gives lifelong protection) (Fig. 1e). Low to moderate levels of immunization, however, will greatly help to control mortality and morbidity, which is of greatest significance in the 0–5-yr-old age classes. Care must, however, be exercised to prevent an associated build-up of susceptible individuals in teenage and adult age classes; at present, such individuals have acquired a degree of immunity to infection from repeated exposure in childhood. Once the properties of a potential vaccine are more clearly defined, mathematical studies can help to quantify such risks. In the case of helminth infections, vaccines will probably only provide protection of short duration against re-infection, and

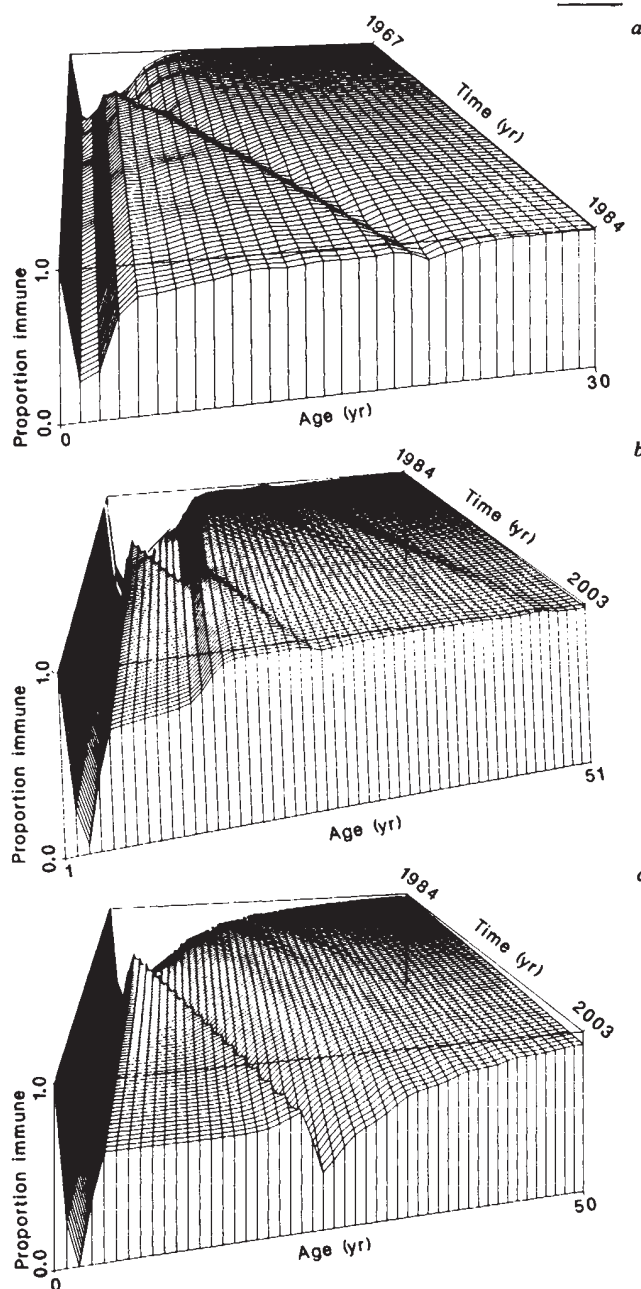


Fig. 4 *a*, Model predictions (equations (1)) of the impact of mass vaccination on the age-serological profiles for measles antibodies (the proportion in each age class who are immune either as a result of maternally derived protection (in the infants), recovery from natural infection, or vaccination) in England and Wales²⁵. Mass vaccination was introduced in year 1 (year 0 denotes the pre-vaccination profile, based on the age-specific forces of infection defined in Fig. 3a) and maintained for 16 yr with the age-specific vaccination rates recorded in England and Wales (mean age at vaccination ~2.2 yr; ~50–60% of each yearly cohort was immunized by age 5 yr). The age axis denotes the intervals 0, 0.5, 1.0, 2.5, ... etc. yr. Note how mass vaccination (in the age range 1–3 yr) acts to decrease the proportion seropositive in the older age classes (by reducing the overall force of infection within the community below the level in the pre-vaccination community). The valley of seronegativity running diagonally across the surface denotes the ageing of the last cohorts of young children to be excluded from mass immunization (see ref. 25 for further details). *b*, *c*, Similar to *a* but denoting model predictions of the impact of vaccination on the age-serological profiles for rubella antibodies over the period 1984–2030 in the United Kingdom in the male (*c*) and female (*b*) segments of the population. The basis of the predictions is described in ref. 67. In the period 1970–84 a single-stage UK policy was in operation (vaccination of teenage girls between the ages of 10–15 yr, with the observed vaccination rates for England and Wales). In 1985 a two-stage (UK-US) policy was begun (vaccination of teenage girls at 1984 rates, vaccination of 60% of 2-yr-old boys and girls) to be maintained up to 2030. Note how the introduction of the two-stage policy alters the level of herd immunity in both the male and female segments of the population. The age axis records the intervals 0, 0.5, 1.5, 2.5, ... 49.5 yr.

hence the creation of herd immunity may require repeated cohort immunization. Cost factors and other practical issues are therefore likely to be of great importance in developing countries when considering the benefits of mass vaccination versus mass chemotherapy.

Conclusions

Many difficulties surround the attainment of sufficient levels of herd immunity to eradicate common infections in developed and developing countries. Theory can define the level of vaccination coverage required for elimination, but success in practice depends on economic and motivational issues. Vaccines against protozoan and helminth parasites will undoubtedly be of importance in the future (certainly to the Western traveller to tropical regions, but less certainly to those who live in these areas), and research towards their development must remain a priority. This emphasis, however, should not distract attention from the importance of epidemiological research. Effective community-

wide control of many of the world's major infections is likely to depend on integrated approaches, involving some combination of selective or targeted vaccination and chemotherapy (directed towards those most predisposed to infection and morbidity), vector control and efforts to improve nutrition, hygiene, sanitation and education. Insights founded on an understanding of the population dynamics of infection will, therefore, be of great importance in any quantitative assessment of control policy. The recent convergence of mathematical theory and observation in epidemiology has created a powerful set of tools for the design and evaluation of community-based programmes of disease control, provided they are used sensibly. At present the potential value of these techniques is not widely appreciated.

We thank the Rockefeller Foundation (R.M.A. and R.M.M.), the Department of Health and Social Security, UK (R.M.A.), and the NSF (grant BSR83-03772 to R.M.M.) for financial support.

1. *Final Report of the Global Commission for the Certification of Smallpox Eradication* (WHO, Geneva, 1980).
2. Katz, S. L. *Pediatrics* **71**, 653-654 (1983).
3. Centres for Disease Control. *Morbidity Mortality Weekly Rep.* **32**, 1-125 (1983).
4. Noah, N. D. *Br. med. J.* **289**, 1476 (1984).
5. Brook, C. G. D. *Br. med. J.* **286**, 1082-1083 (1983).
6. Editorial *Lancet* **ii**, 1343-1344 (1983).
7. Campbell, A. G. M. *Archs Dis. Childh.* **58**, 3-5 (1983).
8. Hinman, A. R. *Rev. infect. Dis.* **4**, 933-936 (1982).
9. Walsh, J. A. & Warren, K. S. *New Engl. J. Med.* **301**, 967-974 (1979).
10. Creese, A. L. & Henderson, R. H. *Bull. Wild Hlth Org.* **58**, 491-497 (1980).
11. Henderson, R. H. *Wild Hlth February/March* 32-35 (1979).
12. De Quadros, C. A. *Proc. R. Soc. B208*, 111-118 (1980).
13. Jilg, W. et al. *Lancet* **ii**, 1174-1175 (1984).
14. Walgate, R. *Nature* **312**, 190 (1984).
15. Weatherall, D. *Lancet* **ii**, 1440-1443 (1984).
16. Mitchison, N. A. *Nature* **308**, 112-113 (1984).
17. *Bull. Wild Hlth Org.* **61**, 81-92 (1983).
18. Maddox, J. *Nature* **310**, 541 (1984).
19. Brown, K. *Nature* **311**, 300 (1984).
20. Editorial *Lancet* **i**, 1053-1055 (1984).
21. Popovic, M., Sornghadharan, M. G., Read, E. & Gallo, R. C. *Science* **224**, 497-500 (1984).
22. Walsh, J. A. *Rev. infect. Dis.* **5**, 330-340 (1983).
23. Anderson, R. M. & May, R. M. *Science* **215**, 1053-1060 (1982).
24. Anderson, R. M. & May, R. M. *J. Hyg., Camb.* **90**, 259-325 (1983).
25. Anderson, R. M. & May, R. M. *J. Hyg., Camb.* **94**, 365-436 (1985).
26. Anderson, R. M. & May, R. M. *IMA J. math. Appl. Biol. Med.* **1**, 233-266 (1984).
27. Dietz, K. *Lect. Notes Biomath.* **11**, 1-15 (1976).
28. Hethcote, H. W. *Am. J. Epidemiol.* **117**, 2-13 (1983).
29. Hoppenssteadt, H. C. *J. Franklin Inst.* **297**, 325-333 (1974).
30. Schenzle, D. J. *IMA J. math. Appl. Biol. Med.* **1**, 169-191 (1984).
31. Anderson, R. M. & May, R. M. *Nature* **280**, 361-367 (1979).
32. May, R. M. & Anderson, R. M. *Nature* **280**, 455-461 (1979).
33. Anderson, R. M. in *Population Biology of Infectious Diseases* (eds Anderson, R. M. & May, R. M.) 149-176 (Springer, Berlin, 1982).
34. Anderson, R. M. (ed.) *Population Dynamics of Infectious Diseases* (Chapman & Hall, London, 1982).
35. Dietz, K. in *Epidemiology* (eds Ludwig, D. & Cooke, K. L.) 104-121 (Society for Industrial and Applied Mathematics, Philadelphia, 1975).
36. Ross, R. *Br. med. J.* **i**, 546-547 (1915).
37. Macdonald, G. *The Epidemiology and Control of Malaria* (London, University Press, 1957).
38. Cvjetanovic, B., Grab, B. & Uemura, K. *Bull. Wild Hlth Org.* **56**, 1-143 (1978).
39. Anderson, R. M. & May, R. M. *Nature* **297**, 557-563 (1982).
40. Anderson, R. M. & May, R. M. *Adv. Parasit.* **25**, 1-101 (1985).
41. Bartlett, M. S. *Proc. 3rd Berkeley Sump. Math. statist. Probabilities* **4**, 81-109 (1956).
42. Cliff, A. D., Haggett, P., Ord, J. K. & Versey, G. R. *Spatial Diffusion: A Historical Geography of Epidemics in an Island Community* (Cambridge University Press, 1981).
43. Bartlett, M. S. *J. R. statist. Soc. A120*, 48-70 (1957).
44. Bartlett, M. S. *Stochastic Population Models in Ecology and Epidemiology* (Methuen, London, 1960).
45. Hethcote, H., Stech, H. W. & Van den Driessche, P. *SIAM J. appl. Math.* **40**, 1-9 (1981).
46. Schwartz, I. B. & Smith, H. L. *J. Math. Biol.* **18**, 233-254 (1983).
47. London, W. P. & Yorke, J. A. *Am. J. Epidemiol.* **98**, 453-468 (1973).
48. Yorke, J. A. & London, W. P. *Am. J. Epidemiol.* **98**, 469-482 (1973).
49. Yorke, J. A., Nathanson, N., Pianigiani, G. & Martin, J. *Am. J. Epidemiol.* **109**, 103-123 (1979).
50. Aron, J. L. & Schwartz, I. B. *J. theor. Biol.* **110**, 665-679 (1984).
51. Anderson, R. M., Grenfell, B. T. & May, R. M. *J. Hyg., Camb.* **93**, 587-608 (1984).
52. May, R. M. & Anderson, R. M. *Math. Biosci.* (in the press).
53. Smith, C. E. *Proc. R. Soc. Med.* **63**, 1181-1190 (1970).
54. Albrecht, P., Ennis, F. A., Saltzstein, E. J. & Krugman, S. *J. Pediatr.* **91**, 715-718 (1978).
55. Heyman, D. L., Mayben, G. K., Murphy, K. R., Guyer, B. & Foster, S. O. *Lancet* **ii**, 1470-1471 (1983).
56. *Expanded Programme for Immunisation: Programme Statement* (WHO, Geneva, 1976).
57. Editorial *Lancet* **i**, 438-439 (1985).
58. Risi, J. B. *Assign. Child* **65/68**, 137-157 (1984).
59. Warren, K. S. & Mahmoud, A. A. F. *Tropical and Geographical Medicine* (McGraw-Hill, New York, 1984).
60. Breman, J. G. & Arita, I. *New Engl. J. Med.* **303**, 1263-1273 (1980).
61. Anderson, R. M. & May, R. M. *New Scientist* **96**, 410-415 (1982).
62. Fine, P. E. M. & Clarkson, J. E. *Int. J. Epidemiol.* **11**, 15-25 (1982).
63. Fine, P. E. M. & Clarkson, J. E. *J. Hyg., Camb.* **92**, 21-36 (1984).
64. Gregg, N. M. *Trans. ophthal. Soc. Aust.* **3**, 35-36 (1941).
65. Hansow, J. B. & Dudgeon, J. A. *Viral Diseases of the Foetus and Newborn* (Saunders, London, 1978).
66. Brunell, P. A. in *Pediatric Infectious Diseases* (eds Feigin, R. D. & Cherry, J. D.) 123-1235 (Saunders, London, 1981).
67. Anderson, R. M. & Grenfell, B. T. *J. Hyg., Camb.* (in the press).
68. Sydenstricker, E. & Hedrick, A. W. *US publ. Hlth Rep.* **44**, 1537-1543 (1928).
69. Fine, P. E. M. & Clarkson, J. A. *Int. J. Epidemiol.* **11**, 5-14 (1982).
70. Black, F. J. *Dis. Childh.* **103**, 242-249 (1962).
71. Kermack, W. D. & McKendrick, A. G. *Proc. R. Soc. A115*, 13-23 (1927).
72. Hethcote, H. W. *Theor. Popul. Biol.* **14**, 338-349 (1978).
73. Becker, N. *Biometrics* **35**, 295-305 (1979).
74. May, R. M. & Anderson, R. M. *Math. Biosci.* **72**, 83-111 (1984).
75. Schenzle, D. *Lect. Notes Biomath.* **57**, 171-178 (1985).
76. Lajmanovich, A. & Yorke, J. A. *Math. Biosci.* **28**, 221-236 (1976).
77. Post, N. M., DeAngelis, D. L. & Travis, C. C. *Math. Biosci.* **63**, 289-302 (1983).
78. Bodmer, W. J. *R. Coll. Phys., Lond.* **14**, 43-50 (1980).
79. Brady, J. A., Sever, J. L., McAlister, R., Schiff, G. M. & Cutting, R. *J. Am. med. Ass.* **191**, 619-626 (1965).
80. Christensen, P. E. et al. *Danish med. Bull.* **1**, 2-6 (1953).
81. Macgregor, J. D., Macdonald, J., Ingram, E. A., McDonnell, M. & Marshall, B. *Br. med. J.* **282**, 434-436 (1981).
82. Pandey, J. P., Zollinger, W. D., Fudenberg, H. H. & Loadholt, C. B. *Clin. Invest.* **68**, 1378-1380 (1981).
83. Spencer, M. J. et al. *Immunogenetics* **4**, 365-372 (1977).
84. Kato, S., Muranaka, S., Takakura, I., Kimura, M. & Tsum, K. *Antigen* **19**, 140-145 (1982).
85. Hoeking, C. S. & Robertson, D. M. *Clinics Immun. Allergy* **1**, 631-639 (1981).
86. Barbour, A. D. *Trans. R. Soc. trop. Med. Hyg.* **72**, 6-15 (1978).
87. Dietz, K. *Lect. Notes Biomath.* **39**, 264-277 (1980).
88. Hethcote, H. W. & Yorke, J. A. *Lect. Notes Biomath.* **56**, 1-105 (1984).
89. Conrad, G. L., Klevis, G. S., Rush, B. & Darrow, W. W. *Sex. Transmitted Dis.* **8**, 241-244 (1981).
90. Rothenburg, R. B. *Am. J. Epidemiol.* **117**, 688-694 (1983).
91. Bloom, A. L. *Lancet* **i**, 1452-1455 (1984).
92. Safai, B. et al. *Lancet* **i**, 1438-1440 (1984).
93. Goedert, J. J. et al. *Lancet* **ii**, 711-715 (1984).
94. Fine, P. E. M. & Clarkson, J. A. *Int. J. Epidemiol.* **11**, 5-14 (1982).
95. Morley, D. *Proc. R. Soc. Med.* **57**, 846-849 (1969).
96. Black, F. L. in *Viral diseases of South East Asia and Western Pacific* (ed. Mackenzie, J.) 297-307 (Academic, Sydney, 1982).
97. Heyman, D. L. et al. *Br. med. J.* **285**, 531-533 (1982).
98. Kok, P. W., Kenya, P. R. & Ensering, H. *Trans. R. Soc. trop. Med. Hyg.* **77**, 171-176 (1983).
99. McLean, A. *Lect. Notes Biomath.* (in the press).
100. Newark, P. *Nature* **311**, 510 (1984).
101. McAleer, W. J. et al. *Nature* **307**, 178-180 (1984).
102. Francis, D. P. *Rev. infect. Dis.* **5**, 321-329 (1983).
103. Stevens, C. E., Beasley, R. P., Tsui, J. & Lee, W. C. *New Engl. J. Med.* **292**, 771-774 (1975).
104. Maupas, P. et al. *Lancet* **i**, 289-292 (1981).
105. Coppel, R. L. et al. *Nature* **310**, 789-792 (1984).
106. Phillip, M., Parkhouse, R. M. E. & Ogilvie, B. M. *Nature* **287**, 538-540 (1980).
107. Cross, G. A. M. *Proc. R. Soc. B202*, 55-68 (1978).
108. Turner, M. J. *Parasitology* **88**, 613-621 (1984).
109. Deloron, P., Jaureguierry, G., LeBras, J. & Coulanges, D. *Lancet* **i**, 444-446 (1985).
110. Warren, K. S. *J. infect. Dis.* **127**, 595-561 (1978).
111. Parkhouse, R. M. E. (ed.) *Parasitology* **88** (1984).
112. Schad, G. A. & Anderson, R. M. *Science* **228**, 1537-1540 (1985).
113. Anderson, R. M. & May, R. M. *Nature* **315**, 439-496 (1985).
114. Grenfell, B. T. & Anderson, R. M. *J. Hyg., Camb.* (in the press).
115. *Morbidity and Mortality Weekly Report* No. 30, 362-363 (Centers for Disease Control, Washington, 1981).
120. Clarke, M., Schild, G. C., Boursted, J., McGregor, I. A. & Williams, K. *Bull. Wild Hlth Org.* **58**, 931-935 (1980).
117. Black, F. L. *J. Immun.* **83**, 74-83 (1959).
118. Shah, U., Banerji, K. L., Nanavati, A. N. D. & Mehta, N. A. *Bull. Wild Hlth Org.* **46**, 130-138 (1972).
119. Molineaux, L. & Gramiccia, G. *The Garki Project* (WHO, Geneva, 1980).
121. Maupas, P. & Melnick, J. L. *Prog. med. Virol.* **27**, 1-5 (1981).
120. Clarke, M., Schild, G. C., Boursted, J., McGregor, I. A. & Williams, K. *Bull. Wild Hlth Org.* **58**, 931-935 (1980).
121. Lundstrom, R., Svedmyr, A., Hagbard, L. & Kaijser, K. *Acta paediat. scand.* **56**, 279-285 (1967).
122. Mortimer, P. P. *Brit. med. J.* **2**, 1523-1524 (1978).
123. Anderson, R. M. *J. Hyg., Camb.* **89**, 1-8 (1982).

Continental rifts as a setting for regional metamorphism

Stephen M. Wickham* & E. Ronald Oxburgh

Department of Earth Sciences, University of Cambridge, Cambridge CB2 3EQ, UK

During very high-temperature/low-pressure Hercynian metamorphism in the Pyrenees the crust began to melt at ~ 12 km and stable isotopes show that it was flushed by circulating seawater to that depth. There is no evidence for crustal collision and the tectonic setting for this, and maybe all high-temperature/low-pressure metamorphism, is a zone of continental rifting.

RIFT valleys are familiar features of the Earth's surface. Some, such as the Rhine Graben, are active today and have associated earthquakes and volcanic activity; others, such as the Midland Valley of Scotland, may be recognized only by the presence of a down-faulted belt of younger rocks preserved in a more ancient terrane. Such rifts must have some expression deeper in the crust, but, at present, its nature is unknown, and there is no way of recognizing them in ancient and deeply eroded areas. Here we link continental rifting in a genetic association with high-temperature/low-pressure (high- T /low- P) regional metamorphism, an equally common crustal feature but one of uncertain tectonic setting. We argue that the transient physical conditions required for the metamorphism can probably be attained only during rifting; and that the geophysical signatures characteristic of the crust below an active rift can best be satisfied by rocks with the physical properties of those undergoing very high grade metamorphism. We support these conclusions with a detailed study of an exceptionally well exposed Hercynian metamorphic terrain in the eastern Pyrenees.

Eastern Pyrenees

In the Pyrenees, regional early Tertiary uplift has exposed an extensive Hercynian basement terrain (Fig. 1) including abundant granitoid intrusives and several condensed metamorphic sequences that imply exceptionally high thermal gradients^{1,2}. Subsequent Alpine metamorphism and deformation has been slight and is confined to the main fault zones. In between these faults, relatively pristine slices of basement have been differentially elevated and, in consequence, well-preserved Hercynian rocks from a variety of structural levels are often clearly exposed in high mountainous terrains. These range from Upper Carboniferous/Lower Permian sedimentary strata and volcanics,

through Lower Palaeozoic metasediments to granulites from deep crustal levels.

The Trois Seigneurs Massif (Fig. 1) is a thrust slice of Hercynian basement occurring to the north of the zone of maximum uplift. Here, a complete section from fossiliferous Silurian shales through andalusite and sillimanite schists to migmatites and a heterogeneous S-type granitoid is clearly exposed. A section through part of the metamorphic sequence is shown in Fig. 2. The lithologies in the sequence are dominantly pelitic but rare, thin carbonate bands are also present and provide a complementary control on metamorphic grade. A variety of lines of petrological evidence³ indicate that the onset of melting in the field sequence (believed to correspond to the first appearance of migmatites) was at $\sim 700^\circ\text{C}$ and at 3.0–3.5 kbar: this conclusion is supported by field and petrographic evidence (see below) and by experimental melting studies on the pelitics carried out over a wide P/T range, with various water activities. Metamorphism and anatexis occurred under water-rich conditions with $a_{\text{H}_2\text{O}}$ buffered externally near unity, leading to the generation of large quantities of granitic melt from the pelitic rocks³.

Metamorphism and melting were accompanied by deformation and the formation of flat-lying isoclinal folds with axial surfaces parallel to the isograds in all the higher grade rocks^{1,4,5}, but this deformation had mostly ceased before the metamorphic peak was reached because some of the prograde minerals overgrew these structures. Synmetamorphic leucogranite pods generated by partial melting of pelitic metasediment are commonly boudinaged and extended parallel to the isograds. Both of these extensional deformations have previously been attributed to doming by granitic diapirs⁴ but the boudinage (and possibly other deformational features) could equally well result from regional ductile extension of the metamorphic and granitic rocks. Similar metamorphic sequences are exposed elsewhere in the Pyrenees (for example, in the St Barthelemy Massif^{6,7}).

* Present address: Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA.

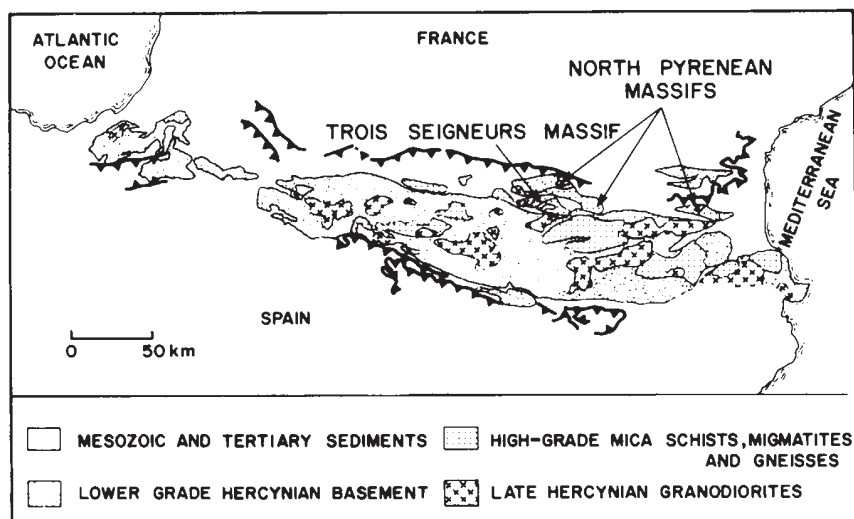
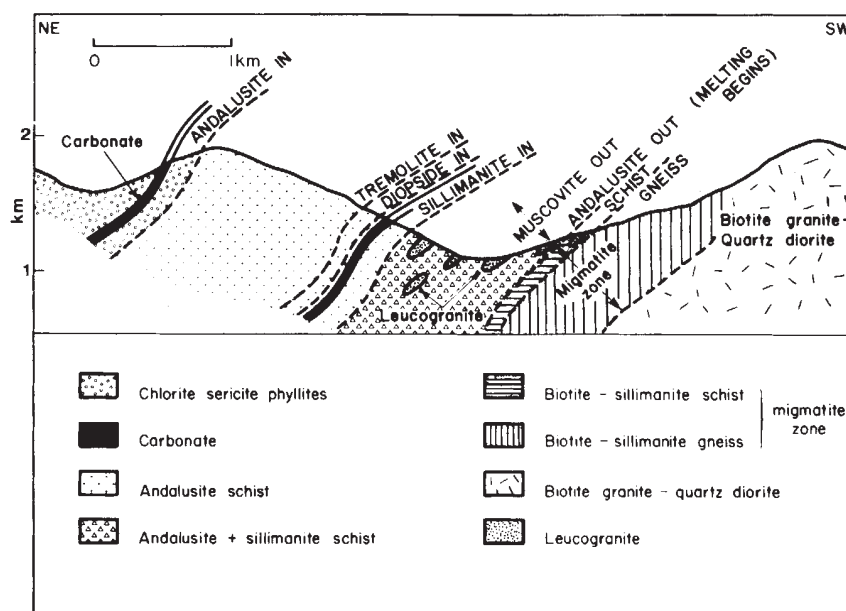


Fig. 1 Map showing the extent of Hercynian basement outcrop in the Pyrenees. The locations of the Trois Seigneurs Massif and other North Pyrenean Massifs are shown.

Fig. 2 Simplified cross-section through the Trois Seigneurs Massif showing the gradational transition from chlorite grade phyllites, through andalusite- and sillimanite-bearing mica schists to migmatites and a per-aluminous granitoid derived by partial melting and homogenization of the pelitic rocks. Note the proximity of low- and high-grade rocks.



Detailed mapping has established that little penetrative deformation has affected the metamorphic sequence since the Hercynian, and it is possible to estimate a thermal gradient for the upper part of the Hercynian crust by measuring the 'stratigraphic distance' (Fig. 2) across a series of isograd surfaces to which temperatures have been assigned. Absolute pressure for the onset of melting can independently be fixed at 3–3.5 kbar by a variety of constraints³, including the mineralogy of the metamorphic sequence (andalusite-sillimanite facies series, staurolite absent, garnet absent except at very Mn-rich compositions), the sequence of isograds (sillimanite in—muscovite out—melting begins), the fact that biotite is a stable phase throughout melting, and the use of biotite-cordierite geobarometry⁸. Having fixed these pressure limits, water activity is also constrained to be high because of the isograd sequence and because biotite persists throughout melting (see refs 3, 9). Where the bases of the high-grade metamorphic sequences are exposed they are seen to pass downwards through migmatite zones into anatexic granites containing disorientated fragments of the more refractory constituents of the former country rock.

The metamorphic reactions corresponding to the isograds documented in the field give a P/T array with a slope of $80\text{--}100\text{ }^{\circ}\text{C km}^{-1}$ for the Hercynian upper crust during metamorphism in the Trois Seigneurs Massif (Fig. 3). Similar metamorphic sequences in many other parts of the Pyrenees^{1,10,11} imply that exceptionally high gradients were developed on a regional scale. Note that slopes of such P/T arrays do not necessarily correspond to true crustal thermal gradients, insofar as the conditions corresponding to the P/T points in the array may have been attained at different times [see ref. 12]. Without a proper understanding of the processes controlling heat and cooling, the significance of such an array is ambiguous.

Hercynian crust during metamorphism

If the P/T array shown in Fig. 3 corresponds to a temperature gradient that existed within the upper part of the Hercynian crust, it could not have continued to depths ≥ 14 km without almost total melting of the crust below this depth, regardless of lithological composition or water content. A similar conclusion is reached by considering the average thermal gradient for the upper crust derived by joining any point in the array to a mean value for the surface temperature.

The crustal temperature distribution implied by these observations cannot represent a steady state: both the geological and geobarometric evidence indicate that the exceptionally high temperatures were not attained by burial, thereby leaving several alternatives. Heat could have been advected to the upper crust

by magma or other fluids from, say, the lower crust or the mantle^{13,14}. A mafic magma could, if present in sufficient quantity and emplaced at a suitable depth (say, 15 km), maintain for a limited period the conditions observed. Such an intrusion would, however, have to be a significant fraction of the crustal thickness (~ 10 km).

Alternatively, high temperatures in the upper crust could have been brought about by a change in the temperature at the base of the crust and thus an increased crustal conductive heat flow. An increase in the temperature at the base of the crust sufficient to give the observed upper crustal conditions, would inevitably also be high enough to melt the lower crust (for any reasonable composition, wet or dry) and in this case heat would be transferred advectively by the ascent of granitoid plutons [see for example, refs 4, 15].

Other evidence

Four other lines of evidence constrain the behaviour of the crust in the eastern Pyrenees during the metamorphic and deformational processes described above.

(1) Extensive granodioritic bodies (such as Mont Louis, Maladeta, and Querigut) have penetrated the whole area; they cut all other lithologies described above and develop contact metamorphic aureoles against them. They were, however, probably emplaced shortly after the cooling of the metamorphic

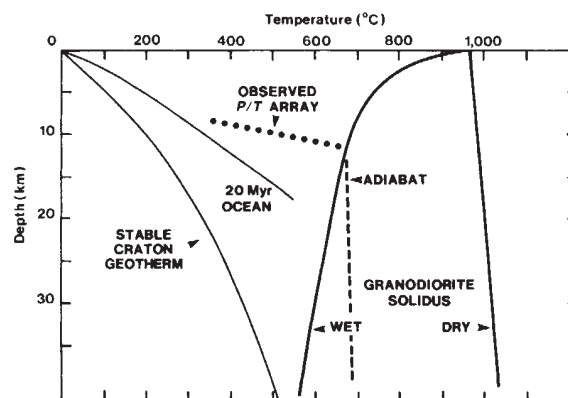


Fig. 3 Observed Trois Seigneurs P/T array compared with typical geotherms for a stable craton and 20-Myr-old oceanic crust. Anatexis of pelitic metasediment in the Trois Seigneurs migmatite zone occurred at $700\text{ }^{\circ}\text{C}$ and 3.0–3.5 kbar.

sequences (see below) and their isotopic and bulk chemical properties suggest that they were derived from the deeper crust^{16,17}. This shows that partial fusion of the Hercynian crust was not restricted to the anatectic zones described earlier. It is, therefore, possible, although on the grounds of timing not likely, that the ascent of the granodioritic magmas transported heat for upper crustal anatexis.

(2) There is local evidence for the incorporation of mafic material in the main, high level, anatectic zone. At higher structural levels mafics are virtually unknown. If the upper crust was heated by the intrusion of mafic magmas, they must mostly have ponded in the middle crust and risen little above that level.

(3) In various places in the eastern Pyrenees, fault-bounded slices of granulite and amphibolite facies gneisses are exposed (see refs 14, 18). The age of these bodies is not well established but in all probability their main metamorphic mineral assemblages are Hercynian, and in any case not younger. Different slices show equilibration temperatures around 700–850 °C and 3.5–7 kbar (ref. 14). Many of them also show evidence of small scale partial melting with development of leucosomes. If these rocks are Hercynian in age they show that the 'middle' crust (for a conventional crustal thickness) was close to its melting temperature at that time. The slices were emplaced to their present levels as a result of Tertiary deformation.

(4) Whereas the previous three considerations relate to the middle and deeper part of the crust during metamorphism, stable isotope studies can be used to constrain processes nearer the surface. Oxygen isotope studies were carried out on 95 rocks and minerals at different levels in the metamorphic sequence. It was shown¹⁹ that in the Trois Seigneurs Massif, the isotopic composition of oxygen at all exposed levels deeper than the andalusite-in isograd has been homogenized at $\delta^{18}\text{O}$ values of +11 to +13‰. Original values in pelitic lithologies would have been about +15‰. Material balance calculations suggest that isotopic compositions were homogenized by contact with a circulating volume of aqueous fluid with total mass probably >30% of the mass of rock affected¹⁹. This means that the fluid is almost certainly of external origin. Deuterium/hydrogen studies on the same material indicate that the fluid was rich in D with a δD of –10 to –15‰. This corresponds to a probable value for late Carboniferous seawater. These considerations suggest, therefore, that the metamorphic-anatectic sequence and the overlying sediments were flushed by circulating seawater during or before the metamorphic maximum.

Pyrenees in the late Palaeozoic

We now examine some broader aspects of the geology that relate to the regional setting of the Hercynian metamorphism in the Pyrenees to see to what extent they are compatible with rifting.

Within the Pyrenees the Palaeozoic sequence that subsequently underwent Hercynian metamorphism extends from the Cambrian to various horizons in the Carboniferous and comprised shales (dominantly), impure sandstones, carbonates and locally quartzites. There is relatively good palaeontological control of ages and the Devonian and Carboniferous are effectively unmetamorphosed. Conversely the highest grades of metamorphism are restricted to the Cambro-Ordovician rocks. The youngest strata in the Trois Seigneurs area are Silurian, but nearby in the Axial Zone the Westphalian (298–315 Myr) is present. Note that the cooling ages of the Trois Seigneurs granodiorite (post-metamorphic) and probably also the metamorphic sequence are 315 ± 10 Myr. This implies that metamorphism of the lower parts of the sedimentary pile was taking place at the same time as deposition was continuing at the surface. The metamorphic maximum was shortly followed by the ascent of granodiorite bodies that show discordant relations with both metamorphosed and unmetamorphosed Palaeozoic. However, the isotopic composition of oxygen in the Trois Seigneurs granodiorite is undisturbed and so the high level hydrothermal circulation that affected the metasedimentary sequence had ceased before its emplacement.

This main sequence of Palaeozoic deposits in the Pyrenees is terminated by an unconformity at which conglomerates, volcanics and continental deposits of Permian or even Stephanian age commonly overlie Westphalian rocks. The time interval represented by the unconformity varies from place to place.

Ophiolitic rocks have not been reported from the Hercynian Pyrenees nor is there any other evidence of involvement of oceanic rock types. There has, however, been substantial Mesozoic and Tertiary movement of Iberia with respect to France and the early Mesozoic position of Iberia cannot be established with confidence. There is even less control on its relative position in the late Palaeozoic, and therefore it is not clear which areas outside the Pyrenees themselves comprised their Palaeozoic regional setting. However, there does seem to be a direct connection between the Carboniferous of the western Pyrenees and that of the eastern Cantabrian mountains of northern Spain. Upper Carboniferous sedimentation is there characterized by the development of locally very thick successions in small but deep basins. Strikingly angular unconformities punctuate the Upper Carboniferous from the Lower Westphalian to the Lower Autunian. The main sedimentary characteristics are high sediment input, very thick successions (up to 6,000 m locally with a total thickness >15,000 m), rapid changes of facies, both vertically and laterally, and indications of contemporaneous gravity driven movement. These features led Reading²⁰ to propose that they represented deposition in extensional regions developed within a strike-slip orogenic belt, movement probably being accommodated on NW-SE and WNW-ESE trending dextral faults, a direction that is compatible with the regional movement of Africa with respect to Europe and North America at that time¹².

Carboniferous rocks are not exposed immediately north of the Pyrenees in the Aquitaine basin, but, in any case, Iberia was probably closer to Brittany and south-west England at that time. Arthaud and Matte²¹ have summarized the evidence that during Carboniferous and Permian time much of southern and central Europe was affected by rifting.

The setting of Hercynian metamorphism in the Pyrenees is poorly constrained. There is no evidence of continental collision or ocean closure. There was, however, contemporaneous rifting outside the area, and possibly strike-slip movement²¹. Metamorphism seems to have occurred towards the end of a prolonged period of basin subsidence and deposition. Both the evidence of deep circulation of seawater through the sedimentary pile and geochronological evidence suggest that the region was below sea level and still receiving sediment while metamorphism was going on at depth.

A rifting setting

The sequence and timing of events documented above is quite different from that recorded from areas such as the Alps and Himalayas where metamorphism has followed crustal collision and tectonic burial. Metamorphism in a collision setting is typically associated with: (1) nappe-type folding and/or thrusting on a regional scale; (2) a delay of some tens of millions of years between collision (and thus the termination of sedimentation) and the metamorphic maximum; (3) the early development of high-pressure mineral parageneses; (4) the common, but not ubiquitous, preservation of lithologies derived from former oceans. None of these features is present in the Hercynian of the Pyrenees.

Active rifts have been studied in many parts of the world^{22,23}. Some, such as the East African Rift and the Rhine Graben seem to be purely extensional features²⁴. Others, such as the Salton Sea and the Dead Sea, are extensional features associated with major strike-slip movements. Typically they are seismically active, topographically low and continue to receive sediment throughout their active lives. Some such as the Central and Viking grabens of the North Sea are localized zones of extension within a regional extensional basin. The rifts commonly have associated volcanism, hydrothermal activity, and high, but variable, surface heat flow.

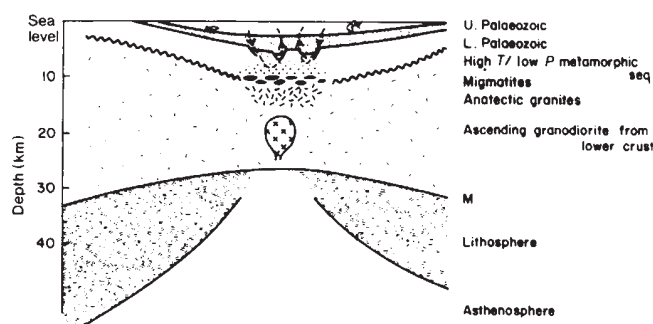


Fig. 4 Schematic tectonic setting for the Hercynian high- T /low- P metamorphism in the Pyrenees. Seawater circulates within the Palaeozoic sedimentary pile, towards the bottom of which metamorphism and large-scale anatexis is occurring at 700 °C and 10–12 km depth. The lower crust is heated to very high temperatures by hot, upwelling asthenosphere, and large bodies of granodiorite magma are generated within this deeper melting zone.

Among the most important geophysical features of presently active rifts are the anomalously low crustal seismic velocities recorded beneath them, and the anomalous attenuation of seismic waves passing through them. Both features may be explained by abnormally high crustal temperatures leading to varying degrees of partial melting. Indeed it is difficult to find any plausible alternative.

Probably the most appropriate modern-day analogue for the processes we propose is to be found in the Salton Sea area of southern California²⁵. The regional tectonic setting is one of strike-slip and extension, associated with the southern continuation of the San Andreas fault system and the extensional regime of the Gulf of California spreading centre. In the latter area, the regional heat flow is $\sim 130 \text{ mW m}^{-2}$ but in the Salton Sea area and some of the associated geothermal fields the heat flow is much higher; temperature gradients of $>150 \text{ }^{\circ}\text{C km}^{-1}$ are measured over the upper 2 km. Locally, there are substantial surface flows of hot water and the drilled sections are, and have been, clearly flushed with hot brines. Muffler and White²⁵ document a series of low-pressure/high-temperature prograde metamorphic reactions from cores in the drilled section.

Although the other rifts, such as the Rhine Graben and the Rio Grande Rift, are associated with high heat flow and local magmatic activity the hydrothermal surface manifestations are less extreme than those observed in the Salton Sea. Seismic reflection work²⁶ indicates the presence of a magma body at 19–20 km depth under Socorro in New Mexico in the southern part of the Rio Grande Rift. There is no way of knowing its temperature or composition but it could be sufficiently hot to

bring about partial melting in the crust immediately above it, providing an analogue for the Trois Seigneurs migmatite. Both P- and S-wave velocities are lower under the Rio Grande Rift than in the regions to either side²⁷.

Discussion and conclusions

The relationship between thermal and mechanical phenomena in rift zones is incompletely understood. It is, however, plausible that the maximum perturbations of crustal temperature should occur where the lithospheric thinning is greatest. Stretching may extend uniformly along a rift or may occur heterogeneously beneath a series of staggered, *en echelon* pull-apart basins. In either case, there may be mafic magmatic activity and the temperature at the base of the crust is elevated as a consequence of lithospheric thinning (Fig. 4); the possible importance of such zones for metamorphic processes has been noted by others^{28–30}. Such zones seem to offer the best possibility for attaining the upper crustal P/T conditions described earlier, and a very hot and partially molten middle and lower crust is suggested both by the attenuation and velocity reduction observed across active rifts today, and by the migmatites and the granulite slices from the Hercynian Pyrenees.

We conclude that the Hercynian 'orogeny' in the Pyrenees was a rifting event, possibly associated with strike-slip motion, that affected a continuously subsiding Palaeozoic Basin; that the deformation and metamorphism occurred well below sea level and that seawater was circulating into the crust to depths of 10 km or more during the process. Such an interpretation is consistent with the regional geology of western Europe at that time.

Rifts probably provide a wide spectrum of crustal environments, but virtually all seem to lead to abnormally high crustal temperatures and thus the possibility of high-temperature/low-pressure metamorphic conditions. Whether seawater flushing occurred in any particular instance, would depend on the degree of local extension (and thus subsidence) and access to the sea. The high-temperature/low-pressure elements in the so-called paired metamorphic belts of the Pacific region probably develop by rifting and extension within island arcs. The tectonic interpretation of high temperature belts in old crustal terrains must be reconsidered if the views presented here are correct.

Some of the work described forms part of a thesis by S.M.W., supported by a NERC research studentship, a research fellowship at Trinity Hall, Cambridge, and both a six-month special graduate studentship and a visiting associateship at the California Institute of Technology. We gratefully acknowledge constructive discussions with Hugh Taylor, Steve Sparks, Mike Bickle, Lee Silver, Robert Hill, Tim Holland, Dave Waters, Jean-Claude Soula, Daniel Vielzeuf and H. J. Zwart. Contribution no. E.S.635 of the Department of Earth Sciences, University of Cambridge.

Received 13 June; accepted 28 August 1985.

- Zwart, H. J. *Leid. Geol. Meded.* **50.1**, 1–74 (1979).
- Fontilles, M. *J. Fac. Sci. Tokyo Univ.* **20**, 181–240 (1981).
- Wickham, S. M. Thesis, Univ. Cambridge (1984).
- Soula, J.-C. *J. struct. Geol.* **4**, 313–342 (1982).
- Verhoef, P. N. W., Vissers, R. L. M. & Zwart, H. J. *Geol. Mijnb.* **63**, 399–410 (1984).
- Zwart, H. J. *Leid. Geol. Meded.* **22**, 419–490 (1959).
- Passchier, C. W. *GUA Pap. Geol.* No. 16 (1982).
- Holdaway, M. J. & Lee, S. M. *Contr. Miner. Petrol.* **63**, 175–198 (1977).
- Thompson, A. B. *Am. J. Sci.* **282**, 1567–1595 (1982).
- Guitard, G. *Le métamorphisme hercynien mésozoïque et les gneiss ocellés du massif du Canigou (Pyrenées orientales)*. Mém. Bur. Rech. Géol. Min. France **63**, 317 (1970).
- Fontilles, M. *Bull. Bur. Rech. Géol. Min. France* **2**, 21–72 (1970).
- Oxburgh, E. R. & England, P. C. *Ecol. Geol. Helv.* **73**, 379–398 (1980).
- Pin, C. & Vielzeuf, D. *Tectonophysics* **93**, 47–74 (1983).
- Vielzeuf, D. Thesis, Univ. Clermont-Ferrand (1984).
- Zwart, H. J. *Geol. Rdsch.* **52**, 38–65 (1962).
- Ben Othman, D., Fourcade, S. & Allègre, C. J. *Earth planet. Sci. Lett.* **69**, 290–300 (1984).

- Vitrac-Michard, A., Albarède, F., Dupuis, C. & Taylor, H. P. Jr. *Contr. Miner. Petrol.* **72**, 57–72 (1980).
- Roux, L. Thesis, Toulouse Univ. (1977).
- Wickham, S. M. & Taylor, H. P. Jr. *Contr. Miner. Petrol.* (in the press).
- Reading, H. G. *9th int. Cong. Sedimentology*, Nice, Theme 4, 289–292 (1975).
- Arthaud, F. & Matte, P. *Bull. geol. Soc. Am.* **88**, 1305–1320 (1977).
- Bridwell, R. J. & Potzick, C. *Tectonophysics* **73**, 15–32 (1981).
- Illies, J. H. *Tectonophysics* **73**, 249–266 (1981).
- Fuchs, K., Bonjer, K. P. & Prodehl, C. *Tectonophysics* **73**, 79–90 (1981).
- Muffler, L. J. P. & White, D. E. *Bull. geol. Soc. Am.* **80**, 157–182 (1969).
- Brown, L. D. et al. in *Rio Grande Rift: Tectonics and Magmatism*, 169–184 (ed. Riecker, R. E.) (American Geophysical Union, Washington DC, 1979).
- Keller, G. R., Braille, L. W. & Schlue, J. W. in *Rio Grande Rift: Tectonics and Magmatism* (ed. Riecker, R. E.) 115–126 (American Geophysical Union, Washington DC, 1979).
- Falvey, D. A. *APEA J.* **14**, 95–106 (1974).
- Dixon, J. E., Fitton, J. G. & Frost, R. T. C. in *Petroleum Geology of the Continental Shelf of North-West Europe* (eds Illing, L. V. & Hobson, G. D.) 121–137 (Heyden, London, 1981).
- Brown, M. & Earle, M. M. *J. metamorphic Geol.* **1**, 183–203 (1983).

Topology of signal recognition particle receptor in endoplasmic reticulum membrane

Leander Lauffer*, Pablo D. Garcia*, Richard N. Harkins†, Lisa Coussens†, Axel Ullrich† & Peter Walter*

* Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143, USA

† Departments of Molecular Biology and Protein Biochemistry, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080, USA

The signal recognition particle (SRP) receptor is an integral membrane protein of the endoplasmic reticulum which, in conjunction with SRP, ensures the correct targeting of nascent secretory proteins to this membrane system. From the complementary DNA sequence we have deduced the complete primary structure of the SRP receptor and established that its amino-terminal region is anchored in the membrane. The anchor fragment and the cytoplasmic fragment contribute jointly to a functionally important region which is highly charged and may function in nucleic acid binding.

We have recently reviewed the function of the signal recognition particle (SRP) and the SRP receptor in the translocation of secretory and lysosomal proteins across the endoplasmic reticulum (ER) membrane, as well as their function in the integration of certain classes of integral membrane proteins into this membrane¹. SRP is thought to recognize the signal peptide on the nascent proteins and arrest their translation in the cytoplasmic space. On interaction of the elongation-arrested ribosome with the ER membrane, which involves a direct interaction of the ribosome-bound SRP with the SRP receptor in the ER membrane, the elongation arrest is released. The SRP receptor is required for protein translocation even in the absence of elongation-arrest², and it thus seems to function primarily to target the SRP-associated nascent polypeptides to the ER membrane. Thereafter, a functional ribosome-membrane junction is established that allows the translocation of the growing polypeptide chain across the ER membrane by a mechanism that is, as yet, poorly understood.

Proteolytic dissection of SRP receptor

Using *in vitro* protein translocation assays as the principal tool, two different approaches³⁻⁸ eventually converged in the discovery and purification of the SRP receptor (or docking protein), as the first—and so far only—integral membrane protein known to play an essential part in this process of protein targeting. Proteolytic dissection of the SRP receptor gave a preliminary picture of its disposition in the ER membrane⁸. The bulk of its mass constitutes a cytoplasmic domain of relative molecular mass (M_r) 52,000 (52K) (see below; by SDS-polyacrylamide gel electrophoresis (PAGE) its size was previously estimated to be 60K) that can be severed from the membrane by a variety of proteases, and retains the ability to reconstitute the activity of proteolysed membranes. By itself, the cytoplasmic fragment—in contrast to the detergent-solubilized intact SRP receptor^{7,8}—is unable to release the SRP-induced arrest of presecretory proteins^{2,6}.

This finding raised the question of whether the 52K SRP receptor fragment still retains affinity for SRP. To directly examine this possibility, we mixed the 52K SRP receptor fragment with a fraction enriched in SRP receptor (Fig. 1, lane 1). We then applied this mixture to an SRP affinity column and monitored the elution of both the 52K SRP receptor fragment and the SRP receptor using Western blot analysis. Intact SRP receptor bound to the SRP affinity column and could be eluted with an increase in the magnesium concentration, as described previously⁹ (Fig. 1, lane 3). In contrast, the 52K fragment was only slightly retarded in the affinity column and was quantitatively recovered in the pooled flow-through and wash fraction (Fig. 1, lane 2). We conclude that the affinity of the 52K receptor fragment for SRP is greatly decreased compared with the intact receptor. Taken together, these findings suggest that the part of

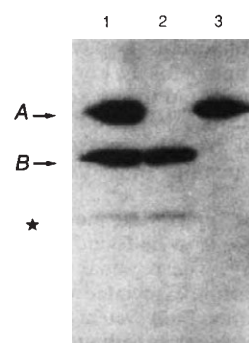


Fig. 1 Chromatography of the SRP receptor and the 52K cytoplasmic fragment on SRP-Sepharose. Lane 1, load fraction; lane 2, combined flow-through and wash fractions; lane 3, column eluate. Arrows indicate the positions of the intact SRP receptor (A) and the 52K cytoplasmic fragment band (B), respectively. The asterisk denotes an additional minor breakdown product (~45K) that was contained in the 52K cytoplasmic fragment fraction.

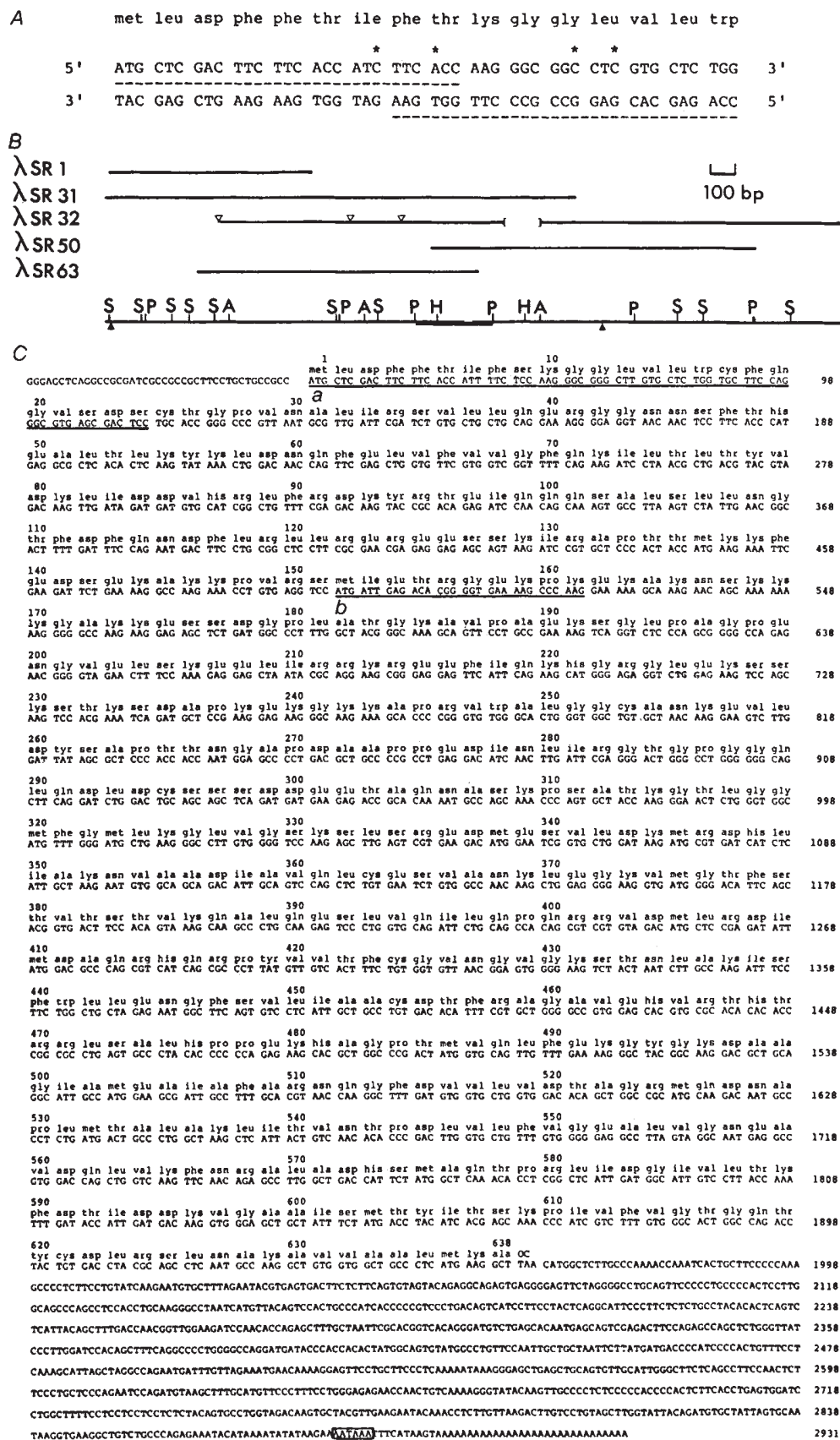
Methods. SRP receptor was partially purified from canine pancreas following the procedure described elsewhere⁹, excluding affinity chromatography on SRP-Sepharose. The 52K cytoplasmic fragment was purified as described elsewhere². Both fractions were mixed to yield 600 μ l of a solution containing 350 ng SRP receptor and 300 ng 52K cytoplasmic fragment in buffer A (50 mM triethanolamine/HOAc pH 7.5, 20 mM KOAc, 5 mM $Mg(OAc)_2$, 250 mM sucrose, 0.5% octaethyleneglycol-mono-*n*-dodecyl ether (Nikkol; Nikko Chem. Corp., Tokyo, Japan), 1 mM dithiothreitol) and 15 mM NaH_2PO_4 . The mixture was applied to a 150- μ l SRP-Sepharose column (containing 40 μ g coupled SRP). The binding capacity of this column was empirically determined to be in excess of 1 μ g SRP receptor bound per 150 μ l of resin. The column was washed with 600 μ l of buffer A containing 50 mM KOAc and eluted with 450 μ l of buffer A containing 10 mM KOAc and 25 mM $Mg(OAc)_2$. Fractions corresponding to 10% of the load were precipitated with trichloroacetic acid and subjected to SDS-PAGE. The gel was blotted onto nitrocellulose. SRP receptor and the 52K cytoplasmic fragment were detected by incubation with a monoclonal antibody against SRP receptor (V. Rath and P.W., to be described elsewhere) followed by incubation with iodinated second antibody and autoradiography²¹.

the SRP receptor that remains membrane-associated after proteolysis does not only contain sequences required to anchor it in the ER membrane, but must also contribute to a functionally important domain of the intact receptor.

Protein sequencing

Clearly, knowledge of the primary sequence of the SRP receptor would help our understanding of its structural organization. We therefore purified SRP receptor and the 52K SRP receptor

Fig. 2 cDNA sequence of the SRP receptor and its predicted amino-acid sequence. **A**, Nucleotide sequence of the probe, based on a preliminary amino-terminal amino-acid sequence of the intact SRP receptor. Two overlapping 27-mers (underlined) were synthesized chemically²³ according to published codon usage frequencies¹⁰. Note that in the preliminary sequence analysis that was used to construct the oligonucleotide probe, amino-acid residue 9 was erroneously assigned as threonine rather than serine (compare with Table 1); this accounts for one of four mismatches (marked by asterisks) of the probe with the cDNA sequence (see **C**). The probe was labelled by first phosphorylating the two oligonucleotides with T4 polynucleotide kinase (PL Biochemicals) and [γ -³²P]ATP (>5,000 Ci mmol⁻¹; Amersham). After annealing the single strands, the complementary sequences were filled-in using *Escherichia coli* DNA polymerase I (Klenow fragment; Boehringer) with [α -³²P]dCTP and [α -³²P]dATP (3,000 Ci mmol⁻¹; Amersham). Poly(A)⁺ mRNA was isolated following standard protocols^{24,25}. Synthesis of cDNA²⁶ cloning into the phage λ gt10 vector²⁷, plaque screening²⁸ and hybridization conditions²⁹ have been described elsewhere. Filters were washed at 42°C with 0.2× SSC. **B**, Schematic representation of the isolated cDNA clones. For screening purposes, purified inserts isolated from agarose gels were randomly primed with a mixed-sequence hexadeoxynucleotide (Pharmacia) and filled-in using the Klenow enzyme and [α -³²P]dATP and [α -³²P]dCTP³⁰, thereby generating probes with a specific activity in excess of 2×10⁸ c.p.m. μ g⁻¹. The 805-bp insert of λ SR1, isolated from a pancreatic cDNA library, was used to screen an MDCK library (see text) and hybridized to λ SR31 and λ SR32, containing, respectively, nucleotides 1–1,866 and 463–2,931 of the complete SRP receptor cDNA sequence (see **C**). λ SR32 is unusual in that it additionally contains three insertions (indicated by triangles; 61 nucleotides (nt) at position 462, 100 nt at position 973, 304 nt at position 1,179) as well as a deletion (indicated by brackets) of 164 nt between positions 1,566 and 1,731. Every one of these aberrations leads to destruction of the open reading frame. The precise nature and possible generation of this clone are presently under investigation. Similar abnormalities have been found in other cDNA clones (A.U. and L.C., unpublished) and are probably due to highly efficient cloning techniques and the use of unfractionated cellular poly(A)⁺ RNA. As SRP receptor sequences seem to be encoded in a single-copy gene (data not shown), we assume that λ SR32 represents an aberrant RNA processing product. Screening of an additional MDCK library with a 304-bp *Pst*I fragment obtained from an m13 clone used to sequence λ SR31 (underlined part of the restriction map) yielded clones λ SR50 and λ SR63, covering positions 1,295–2,555 and 366–1,398, respectively, of the assembled sequence in **C**. Using the m13/dideoxy chain termination method^{31,32}, all five clones were completely sequenced. Restriction sites used for subcloning into the m13 vector are indicated: A, *Av*II; H, *H*indIII; P, *P*stI; S, *S*au3A. All overlapping sequences were identical, except for (1) discrepancies introduced by λ SR32 (see above), (2) nucleotide 1,403 (T in λ SR31 and λ SR63, C in λ SR32; a conservative change) and (3) one of the eight adenine residues around position 550 (λ SR1 and λ SR31) which was found to be deleted in λ SR32 and λ SR63. We attribute these latter discrepancies to misincorporation or slipping, respectively, of reverse transcriptase during the cDNA synthesis. The triangles in the restriction map indicate the beginning and end of the open reading frame. **C**, Nucleotide and predicted amino-acid sequence of the SRP receptor. Linker sequences have been omitted. The amino-terminal sequences obtained by protein sequencing of the intact receptor (**a**) and the cytoplasmic fragment (**b**) are underlined. The region preceding Met 1 translates into Glu (–13)–Leu–Arg–Pro–Arg–Ser–Pro–Pro–Leu–Pro–Ala–Ala–Ala (–1). The polyadenylation signal, AATAAA, is boxed.



fragment from canine pancreas as described previously^{2,7}. Following preparative SDS-PAGE as an additional purification step, the proteins were eluted electrophoretically from the gels and subjected directly to sequence analysis on a gas-phase sequencer. Neither of the polypeptides appeared to be blocked at their amino termini and we were able to determine the first 24 amino acids of the SRP receptor and the first 10 amino acids of the 52K SRP receptor fragment (see Table 1). From these data, we concluded first that the SRP receptor and the 52K SRP receptor fragment have different amino termini, and second that the 52K SRP receptor fragment has a unique sequence which indicates that elastase—the protease used to sever this domain from the membrane—cleaves the SRP receptor in a unique position.

cDNA cloning

Knowledge of the partial primary sequences of these polypeptides allowed us to identify cDNA clones of the SRP receptor. According to codon usage frequencies¹⁰, we designed a single 48-nucleotide(nt)-long double-stranded DNA probe coding for the first 16 amino acids of the receptor (Fig. 2A) and used it to screen a cDNA library in phage λ gt10, constructed by oligo(dT)-priming of canine pancreatic poly(A)⁺ RNA (see Fig. 2 legend). From the library (containing ~500,000 clones), we isolated one recombinant phage (λ SR1, Fig. 2) that gave a strong hybridization signal and which contained an insert of 805 base pairs (bp). Re-screening of the library with this cDNA insert (see Fig. 2 legend for experimental details) did not yield any new cDNA clones, indicating that the isolated clone was the only

Table 1 Sequence analysis of the SRP receptor and the 52K SRP receptor fragment

Cycle	SRP receptor		52K cytoplasmic fragment	
	Amino acid	Yield (pmol)	Amino acid	Yield (pmol)
1	Met	194	Met	52
2	Leu	198	Ile	49
3	Asp	166	Glu	32
4	Phe	135	Thr	Not quantitated
5	Phe	180	Arg	27
6	Thr	48	Gly	31
7	Ile	138	Glu	25
8	Phe	130	Lys	33
9	Ser	22	Pro	11
10	Lys	61	Lys	27
11	Gly	67		
12	Gly	75		
13	Leu	96		
14	Val	88		
15	Leu	80		
16	Trp	12		
17	Not detected			
18	Phe	45		
19	Gln	16		
20	Gly	38		
21	Val	19		
22	Ser	5		
23	Asp	Very low		
24	Ser	4		

Both proteins were purified from canine pancreas as previously described^{2,9}. After preparative PAGE in SDS using a 7–11% gradient gel and electroelution²², the samples were applied directly to sequencer filters precycled with 1.5 mg of Polybrene. Sequence analysis was performed in a gas phase sequencer (Applied Biosystems model 470 A). The phenylthiohydantoin amino acids were separated using a Waters HPLC system equipped with a Rainin Microsorb C8 column (0.46 × 25 cm) heated to 45 °C in a gradient of solvent I [0.25 M NaH₂PO₄ pH 4.6/CH₃CN/H₂O (10/18/72 by volume)] to solvent II [H₂O/CN₃CN (70/30)]. Initial yields were 88% and 40%, average repetitive yields amounted to 94% and 93% for SRP receptor and 52K cytoplasmic fragment, respectively. Protein determinations were performed by densitometric scanning of Coomassie blue stained bands in analytical SDS polyacrylamide gels using bovine serum albumin as a standard.

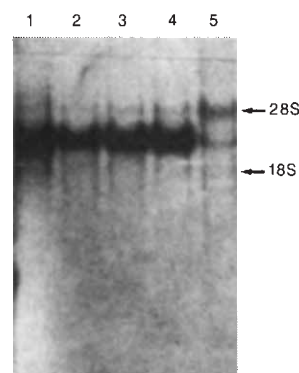


Fig. 3 Northern blot analysis of poly(A)⁺ RNA from MDCK cells (lane 1), canine kidney (lane 2), liver (lane 3), testis (lane 4) and pancreas (lane 5).

Methods. Total cellular RNA was prepared by the guanidine monothiocyanate/LiCl precipitation method²⁴ and selected on oligo(dT)-cellulose²⁵; 3.3 μ g (lanes 1–4) or 6 μ g (lane 5) of RNA was separated in a 1.2% agarose/formaldehyde gel²⁵ and transferred directly to nitrocellulose. Hybridization conditions were exactly as described in Fig. 2 legend using the insert of clone λ SR31 as probe (5×10^5 c.p.m. μ g⁻¹, 1.5×10^6 c.p.m. ml⁻¹). The film was exposed for 3 days. The two additional bands in lane 5 are probably due to nonspecific interaction of the probe with 18S and 28S ribosomal RNA; the relative proportion of these RNA species as visualized by ethidium bromide staining was higher in the pancreatic preparation than in the other lanes.

one containing these SRP receptor sequences. We therefore constructed further cDNA libraries from poly(A)⁺ RNA from MDCK (Madin Darby canine kidney) cells. Screening of these libraries (containing 500,000 recombinant clones) yielded four additional overlapping clones (λ SR31, 32, 50 and 63; schematically described in Fig. 2B) which cover the complete coding and 3'-untranslated sequence of the SRP receptor (see below).

We estimate that the SRP receptor constitutes only 0.1% of the protein contained in a rough ER cell fraction from pancreas. This low abundance, together with the fact that pancreatic messenger RNA largely encodes secretory and not cellular 'household' proteins, could account for the scarcity of SRP receptor sequences in this library. Indeed, in a Northern analysis of poly(A)⁺ RNA isolated from different canine tissues (Fig. 3), the relative abundance of SRP receptor transcripts was substantially higher in tissues with lower secretory activity. The best signal was obtained with RNA from MDCK tissue culture cells, presumably because a rapidly dividing cell is forced to constantly proliferate its ER membrane together with other vital cell functions.

Protein primary sequence

Analysis of the cloned cDNA sequence (2,931 bp) shows a single open reading frame that starts close to the 5' end. This reading frame contains a precise match with the amino-terminal amino-acid sequence of the SRP receptor extending beyond the amino acids used to design the oligonucleotide probe (Table 1, Fig. 2C). The predicted M_r of the encoded protein is 69,684, in good agreement with its apparent M_r of 72K, as estimated previously by SDS-PAGE⁷. Amino-acid analysis of the purified protein yields values close or identical to those derived from its predicted primary structure (data not shown). The cloned DNA extends for 41 nucleotides to the 5' side of the sequence coding for the amino terminus of the SRP receptor. Because this sequence does not contain an in-frame stop codon, we cannot formally exclude the possibility that the SRP receptor could be encoded as a precursor containing a cleavable signal sequence preceding Met 1. However, translation of these nucleotides (see legend to Fig. 2C) does not result in an amino-acid sequence resembling that commonly found in signal peptides¹¹. Therefore, it is possible that—as is the case for other integral membrane proteins (see ref. 12)—the information required for its integra-

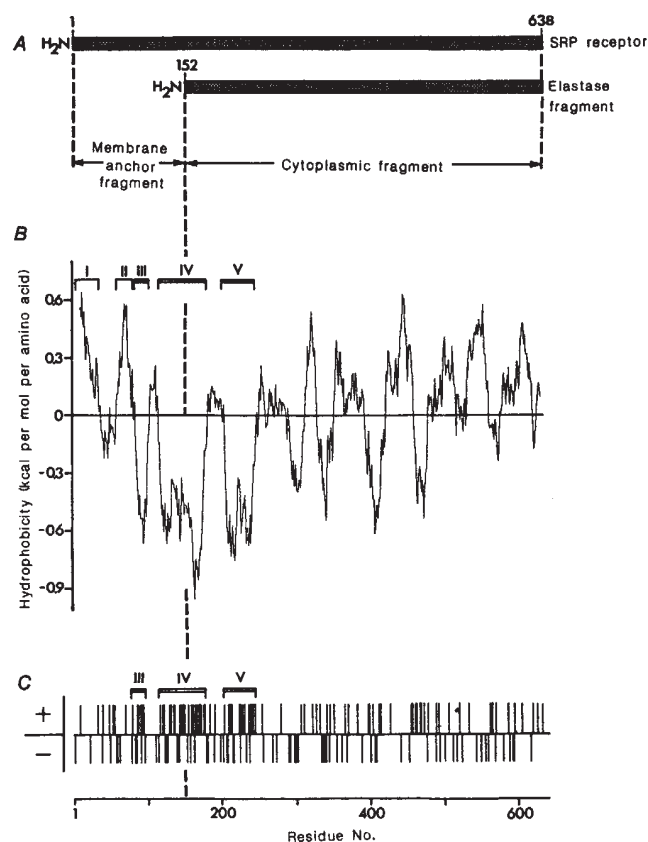


Fig. 4 Physical characteristics predicted from the SRP receptor primary structure. **A**, The locations of the membrane-anchoring fragment and the cytoplasmic domain extending to the carboxy terminus are indicated. Solid boxes mark the regions for which protein sequence data have been obtained; the vertical dashed line indicates the elastase cleavage site. **B**, Hydrophobicity values³³ were averaged for windows of 14 residues and are given as kcal per mol per amino acid of average free energy for the transfer from a hydrophobic to a hydrophilic environment. Regions I and II (amino acids 1–22 and 64–79, respectively) are candidates for sequence stretches interacting with the lipid bilayer. Clusters of charged residues are marked III (amino acids 84–97), IV (129–175) and V (205–243). **C**, Charge distribution along the SRP receptor sequence. All histidine, arginine and lysine residues were assumed to be positively charged (upward-projecting lines) and glutamic and aspartic acid residues to be negatively charged (downward-projecting lines).

tion into the membrane is encoded in an uncleaved signal sequence which in the SRP receptor would be contained in the amino-terminal portion of the protein (see below). The nucleotide context of the presumptive initiating ATG codon (GXXATGY) is also found in other eukaryotic mRNAs¹³, although it diverges from the most commonly found consensus sequence, AXXATGG¹⁴.

The SRP receptor and the 52K cytoplasmic fragment were purified by different methods and monitored using different activity assays^{2,9}. Sequence alignment of the amino-terminal region of the 52K SRP receptor fragment reveals that the fragment begins with Met 152 of the intact receptor; this provides independent confirmation of the identity of the cloned sequence as SRP receptor cDNA. From the position of the amino terminus of the 52K cytoplasmic fragment within the sequence, we conclude that the amino-terminal 151 amino acids (calculated M_r , 17.5K) comprise the membrane-anchoring fragment of the SRP receptor, and the remaining 487 amino acids (calculated M_r , 52K) comprise the cytoplasmic SRP receptor fragment. A schematic representation of the deduced amino-acid sequences of the SRP receptor and the 52K SRP receptor fragment is shown in Fig. 4A.

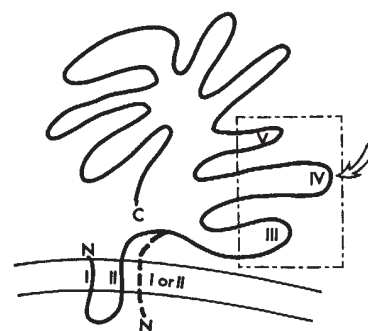


Fig. 5 Model of the disposition of the SRP receptor in the ER membrane. Regions I and/or II are the putative membrane-spanning regions; regions III–V contain the charge clusters described in the text. The arrow depicts the protease-sensitive site.

Analysis of the primary sequence for distribution of hydrophobicity reveals only two distinctly hydrophobic stretches (amino acids 1–22 and 64–79, labelled I and II in Figs 4B, 5) in the amino-terminal part of the sequence. These regions are, in principle, long enough to span the hydrophobic core of the membrane, although they are shorter than most known α -helical transmembrane segments¹⁵. However, either one or both of these regions must be in contact with the hydrophobic core of the lipid bilayer, because the intact SRP receptor can be released from the membrane only after its solubilization with detergents, and—according to the argument outlined above—it is anchored in the membrane through its amino-terminal 151 amino acids.

Apart from these two hydrophobic stretches, the remainder of the membrane-anchoring fragment of the SRP receptor is surprisingly hydrophilic. This character extends across the protease-sensitive site of the intact receptor for about 100 amino acids into the cytoplasmic fragment. In particular, the regions spanning amino acids 84–97, 129–175 and 205–243 show extreme hydrophobicity values (regions labelled III, IV and V in Figs 4B, C and 5) that are due to a remarkable abundance of charged amino acids (both basic and acidic in a relative proportion of ~2:1) which largely appear clustered (Fig. 4C). Notable is the predominantly basic character of these domains, which also dominates the overall charge of the SRP receptor (calculated pI = 9.86).

The remainder of the cytoplasmic portion of the SRP receptor contains periodically spaced regions, some of which exceed in length and hydrophobicity values those which must comprise the membrane anchor segment(s). However, as described above, on proteolysis the 52K SRP receptor fragment is readily released from the membrane, by treatment with high salt concentrations, as a rather stable and 'well-behaved' (that is, non-aggregating) soluble protein. Thus these hydrophobic stretches are probably buried within the protein. There remains the possibility that, during the functional cycle of the SRP receptor, these regions become transiently associated with the hydrophobic core of the lipid bilayer (see below).

Functional implications

When we compared the SRP receptor sequence with other known protein sequences, we found a striking resemblance of the mixed-charge amino-acid clusters (III–V in Fig. 4) to those regions in ribosomal proteins and aminoacyl-tRNA synthetases that are thought to be involved in binding of these proteins to RNA molecules^{16–19}. In addition, when the Dayhoff databank was searched for homologous protein sequences, the best match was found with sea urchin histone H1; there is 30% identity in an overlap extending from residues 120 to 201 of the SRP receptor, primarily due to the frequent occurrence of positively charged amino acids. The finding that regions III–V resemble nucleic acid-binding proteins raises the possibility that the SRP receptor could interact directly (through a domain constituted

by regions III, IV and/or V; indicated as a boxed region in Fig. 5) with the 7SL RNA in SRP, and thus become, albeit transiently, an integral part of SRP. Such a hypothesis is consistent with the fact that the SRP receptor can be eluted from an SRP affinity column (Fig. 1; ref. 9) with a relatively small increase in the magnesium concentration (from 5 to 25 mM), the ionic strength remaining constant. This small concentration difference is not likely to severely affect protein-protein interactions; however, the conformational state of RNA molecules is known to be highly dependent on the concentration of divalent cations²⁰. We therefore consider it likely that the mixed-charge amino-acid clusters III-V constitute or contribute to a domain in the molecule that is involved in its interaction with SRP. If these regions bind to the 7SL RNA molecule in SRP (as suggested by their primary structure), their separation by proteolytic cleavage (arrow in Fig. 5) could account for the apparent lack of activity of both the membrane anchoring fragment and the cytoplasmic fragment of the SRP receptor.

Received 6 August; accepted 30 September 1985.

1. Walter, P., Gilmore, R. & Blobel, G. *Cell* **38**, 5-8 (1984).
2. Siegel, V. & Walter, P. *J. Cell Biol.* **100**, 1913-1921 (1985).
3. Walter, P., Jackson, R. J., Marcus, M. M., Lingappa, V. R. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1795-1799 (1979).
4. Meyer, D. I. & Dobberstein, B. *J. Cell Biol.* **87**, 498-502 (1980).
5. Meyer, D. I. & Dobberstein, B. *J. Cell Biol.* **87**, 503-508 (1980).
6. Gilmore, R., Blobel, G. & Walter, P. *J. Cell Biol.* **95**, 463-469 (1982).
7. Gilmore, R., Walter, P. & Blobel, G. *J. Cell Biol.* **95**, 470-477 (1982).
8. Meyer, D. I., Krause, E. & Dobberstein, B. *Nature* **297**, 647-650 (1982).
9. Gilmore, R. & Blobel, G. *Cell* **35**, 677-685 (1983).
10. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. & Mersier, R. *Nucleic Acids Res.* **9**, 43-74 (1981).
11. Watson, M. E. F. *Nucleic Acids Res.* **12**, 5145-5164 (1984).
12. Kopito, R. R. & Lodish, H. F. *Nature* **316**, 234-238 (1985).
13. Hudson, P., Haley, J., Cronk, M., Shine, J. & Niall, H. *Nature* **291**, 127-131 (1981).
14. Kozak, M. N. *Nature* **308**, 241-246 (1984).
15. Finer-Moore, J. & Stroud, R. M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 155-159 (1984).
16. Ehresmann, B., Reinbolt, J. & Ebel, J. P. *FEBS Lett.* **58**, 106-111 (1975).

One can speculate that in addition to targeting the nascent polypeptide chain to the membrane, the SRP receptor might have a role in the initiation of protein translocation itself. For example, clusters of high charge density positioned in the immediate vicinity of the lipid bilayer, perhaps in conjunction with the hydrophobic stretches in the carboxy-terminal half of the protein, may lead to a local perturbation of the structural integrity of the membrane, which could be used directly or indirectly by the protein translocation machinery.

We thank Dr Peter Seeburg for advice on the DNA sequencing, Dr John Bell for his help with the protein sequencing and Drs D. Agard, B. Alberts, R. Fletterick and R. Stroud for helpful discussions. We acknowledge M. Vasser and P. Jhurane for synthesizing the oligonucleotides. This work was supported by NIH grant GM-32384. L.L. is a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research. P.D.G. receives a scholarship from ODEPLAN-Chile. P.W. is supported by the Chicago Community Trust/Searle Scholars Program.

17. Ehresmann, B., Reinbolt, J., Backendorf, C., Tritsch, D. & Ebel, J. P. *FEBS Lett.* **67**, 316-319 (1976).
18. Putney, S. D. *et al. Science* **213**, 1497-1501 (1981).
19. Walter, P. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 2437-2441 (1983).
20. Cantor, C. R. in *Ribosomes* (eds Chambliss, G. *et al.*) (University Park, Baltimore 1980).
21. Walter, P. & Blobel, G. *J. Cell Biol.* **97**, 1693-1699 (1983).
22. Hunkapiller, M. W., Lujan, E., Ostrander, F. & Hood, L. E. *Meth. Enzym.* **91**, 227-236 (1983).
23. Crea, R. & Horn, T. *Nucleic Acids Res.* **8**, 2331-2348 (1980).
24. Cathala, G. *et al. DNA* **2**, 329-335 (1983).
25. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
26. Gray, A., Dull, T. J. & Ullrich, A. *Nature* **303**, 722-725 (1983).
27. Huynh, T., Young, R. & Davis, R. in *Practical Approaches in Biochemistry* (ed. Grover, D.) (IRL, Oxford, 1984).
28. Benton, W. D. & Davis, R. W. *Science* **196**, 180-182 (1977).
29. Ullrich, A., Berman, C. H., Dull, T. J., Gray, A. & Lee, J. H. *EMBO J.* **3**, 361-364 (1984).
30. Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6-13 (1983).
31. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
32. Schreier, P. H. & Cortese, R. *J. molec. Biol.* **129**, 169-172 (1979).
33. Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105-132 (1982).

Bovine opsin has more than one signal sequence

Martin Friedlander & Günter Blobel

Laboratory of Cell Biology, Rockefeller University, New York, New York 10021, USA

By deletion of selected segments from a bovine opsin complementary DNA clone and subsequent analysis of transcripts in a cell-free translation-translocation system, we have localized two out of four theoretically conceivable signal sequences required for the integration of opsin into microsomal membranes.

THE *in vitro* translation of messenger RNAs for integral membrane proteins in the presence of an appropriate acceptor membrane (representing the *in vivo* site of integration) has proved to be a useful approach for studying the mechanism by which proteins are asymmetrically integrated into membranes.

Experiments of this kind were first carried out using glycoprotein G of the vesicular stomatitis virus as a model integral membrane protein and dog pancreas microsomal membranes as a model acceptor membrane¹⁻³. These experiments suggested that the initial events in the integration of the G protein into microsomal membranes were similar to those that occur in the translocation of secretory proteins. In both cases, the proteins were synthesized with transient amino-terminal signal sequences that were cleaved during or shortly after translocation³. Moreover, the signal sequences of these two proteins were similar and used the same translocation machinery in the microsomal membrane³. Translocation, as well as integration, was shown to be dependent on the signal recognition particle (SRP): translocation (integration) did not occur with KCl-washed (SRP-depleted) microsomal membranes (KRM) and the re-addition of SRP restored this ability (for summary, see ref. 4).

Several proteins that are not synthesized as larger precursors and therefore do not have cleaved signal sequences have been studied using the above assays. Two different integration

mechanisms could be distinguished experimentally⁵⁻¹⁰. In one case, exemplified by cytochrome *b₅*, integration into the lipid bilayer occurs spontaneously and independently of SRP⁵⁻⁸. The segment within the protein that specifies such spontaneous insertion into membranes has been termed an 'insertion' sequence¹¹. In the other case, exemplified by Ca²⁺-ATPase⁸, integration is specified by a signal sequence and is catalysed by SRP.

The present study used bovine opsin as a model for integral membrane proteins that are not synthesized as larger precursors and are polytopic¹¹. Opsin traverses the membrane seven times, resulting in four separate hydrophilic domains being exposed on the exoplasmic side of the microsomal membrane and four on the cytoplasmic side (for summary, see refs 12, 13). We shall call 'trans' domains those portions of the molecule that have been translocated across the membrane such that they extend into the microsomal vesicle lumen (representing the 'outside' of the cell). Those portions that are not translocated across the microsomal membrane and remain in the biosynthetic compartment of the cell are called 'cis' domains. Hypothetical considerations¹¹ indicate that several possible mechanisms could establish this topology. For example, each of the translocated domains may require a separate signal sequence. If this were the case, opsin may contain as many as four uncleaved signal sequences, one for each domain to be translocated. Alternatively, a single

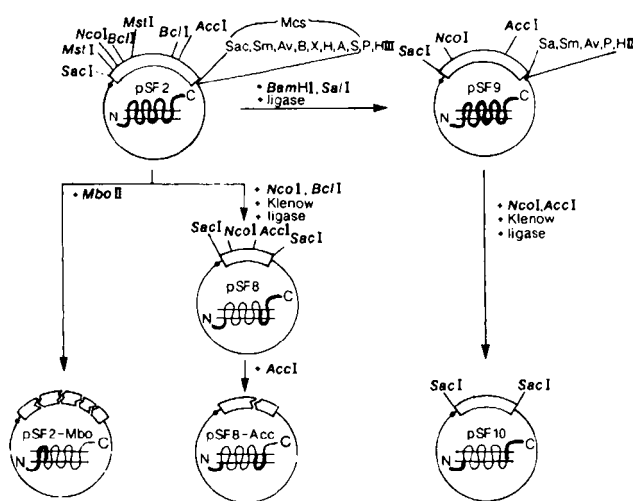


Fig. 1 Construction of opsin-containing SP6 vectors. Mcs, multiple cloning site; Sac, *SacI*; Sm, *SmaI*; Av, *AvuI*; B, *BamHI*; X, *XmaI*; H, *HincII*; A, *AccI*; S, *Sall*; P, *PstI*; HIII, *HindIII*; N, amino-terminus; C, carboxy-terminus; solid circle, SP6 polymerase binding site. A schematic representation of the various forms of opsin encoded by each plasmid is included within each large circle. The two parallel lines represent the membrane lipid bilayer and the opsin molecule is depicted by a waveform that traverses the membrane seven times. The darkened portion of each opsin represents the portion of the molecule encoded by the respective plasmids.

Methods. A cDNA clone (bd20) encoding the entire opsin apoprotein of bovine rhodopsin was obtained from Dr J. Nathans (Stanford University)¹⁴. This 1,639-nucleotide clone was removed from pBR322 by cutting it with *SacI* and ligating the resulting 1,382-base pair (bp) restriction fragment into the *SacI* site of pSF65 (ref. 17). This plasmid (pSF2) was used to transform a *dam*-methylase⁻ strain of *Escherichia coli*, GM48. The insert contained nucleotides 36–1,414 of bd20, 60 nucleotides of the 5'-untranslated region, a coding region of 1,044 nucleotides and 274 3'-untranslated nucleotides. pSF2 was used to prepare various truncation or deletion recombinants. All cloning procedures were as described by Maniatis *et al.*²⁸ and restriction digests were carried out in conditions recommended by New England Biolabs. Recombinant plasmids, as described below, were identified by restriction endonuclease digestion and used to synthesize mRNA¹⁷. Construction of specific plasmids: **pSF2-Mbo**. This truncated form of opsin (encoding amino acids 1–93) was obtained by digesting pSF2 with *Mbo*II. Multiple fragments were generated, but only one 710-bp piece contained the SP6 polymerase binding site continuous with the first 311 nucleotides of the opsin insert. **pSF8**. To obtain this plasmid, pSF2 was digested with *Nco*I and *Bcl*I. The resultant 3' recessed termini were filled-in with Klenow fragment and ligated to produce a recombinant with the correct opsin reading frame, a retained *Nco*I site and no *Bcl*I site. **pSF8-Acc**. This truncated DNA, prepared by digesting pSF8 with *Acc*I, encoded a 79-residue protein. **pSF9**. Identical to pSF2 except that 8 nucleotides of the SP65 multiple cloning site were deleted by digesting pSF2 with *Bam*HI and *Sall* and re-ligating the linearized vector. This was done to remove the *Acc*I site from the multiple cloning site so that the *Acc*I site at nucleotide 961 in the opsin coding region became a unique restriction site. **pSF10**. This plasmid encodes 83 amino acids of opsin, including the 34 amino-terminal residues continuous with, in frame, the carboxy-terminal 48 residues. An additional Cys residue at position 35 is a result of filling-in and ligating. Thus, a form of opsin is obtained which contains none of the transmembrane segments and only the first *trans* and the last *cis* domains. To obtain pSF10, pSF9 was digested with *Nco*I and *Acc*I, filled-in with Klenow and ligated, resulting in the loss of the *Nco*I and *Acc*I sites and the acquisition of an additional *Nla*III site.

signal sequence may be used for the translocation of the NH₂-terminal translocated domain (which is glycosylated at two sites) followed by three insertion sequences, one each for the three remaining translocated domains. Combinations of signal and insertion sequences (one for each translocation domain) are also conceivable. Finally, it is possible that opsin does not contain signal sequences and that its integration occurs exclusively by insertion sequences.

To distinguish between these possibilities and to localize at least some of the 'topogenic' sequences required for integration¹¹, we carried out manipulation at the DNA level, either by truncating a full-length complementary DNA clone of bovine opsin¹⁴ or by cutting the cDNA clone and ligating various segments of interest. The resulting constructs (as well as the uncut full-length cDNA) were then transcribed and the RNAs

	Alkali												IP			
KRM	-	-	+	+	+	+	+	+	+	★	★	-	-	-	+	+
SRP	-	+	-	+	-	+	-	+	-	+	-	-	-	-	+	+
					S	P	S	P	S	P	S	P			E	

The gel image shows 15 lanes. Lanes 1-12 are under 'Alkali' treatment, and lanes 13-15 are under 'IP' treatment. Arrows point to bands labeled 'OP' and 'GLO'. Lane 15 is a control lane.

Fig. 2 Opsin is integrated into microsomal membranes in an SRP-dependent fashion. SP6-generated opsin mRNA¹⁷ and total rabbit reticulocyte RNA were translated in a wheat-germ system with (+) or without (-) SRP (800 U ml⁻¹) or salt-washed membranes (KRM, 2 µl) in a total reaction volume of 24 µl. In lanes 9 and 10, the stars indicate that KRM were present only during extraction with alkali, not during translation. Following translation, samples in lanes 5-12 were treated with alkali (pH 11.0) for 10 min on ice, centrifuged into an alkaline sucrose cushion for 10 min at 30 p.s.i. in a Beckman airfuge and then separated into supernatant (S) and pellet (P) fractions before solubilization and gel loading. The samples in lanes 13 and 14 were immunoprecipitated with antiserum to opsin and then prepared for SDS-PAGE. In lane 15, the sample was treated with Endo H (E) and then immunoprecipitated with anti-opsin serum. OP, primary translation product of opsin; GLO, α and β-globin; IP, immunoprecipitates; alkali, alkali-extracted. Downward-pointing arrows indicate the position of the glycosylated form of opsin.

Methods. Sp6-generated mRNAs were prepared from linearized plasmids (see Fig. 1) as described by Melton and co-workers¹⁷. After transcription, mRNAs were purified by phenol/chloroform extraction and ethanol and then lithium precipitations²⁸. The wheat-germ *in vitro* translation system has been described previously¹⁸; 60 or 100 ng of pSF2 mRNA and 20 A₂₆₀ U of total rabbit reticulocyte RNA were added to each 24-μl reaction mixture. Translations were carried out for 40–60 min at 23 °C and stopped by chilling on ice and precipitation with an equal volume of cold 20% trichloroacetic acid. Preparation of KRM and SRP was as described previously^{29,30}; 2 μl KRM was used with 20 U SRP for each 24-μl translation mixture. Alkali extraction and alkaline sucrose gradient centrifugation were carried out according to Gilmore and Blobel²². Digestion with Endo H was for 12 h as described previously³¹. The SDS-PAGE system has been described elsewhere¹⁶.

translated in a wheat-germ cell-free translation system in the presence or absence of SRP-depleted microsomal membrane and/or the absence or presence of SRP. Our data suggest that opsin contains at least two signal sequences.

Opsin integration requires SRP

Our first objective was to determine whether full-length opsin, which is not synthesized as a larger precursor¹⁵, contains a signal sequence or whether its integration into microsomal membranes¹⁶ is mediated entirely by insertion sequences. A construct of full-length opsin DNA in SP6 (pSF2; see Fig. 1) was transcribed¹⁷ and the RNA translated in a wheat-germ cell-free system¹⁸. Rabbit reticulocyte RNA (coding primarily for the two globin chains) was included in all translations (unless otherwise noted). The synthesized globin served as an internal

control, as its synthesis should not be affected by SRP and it should not be integrated into microsomal membranes. Opsin and the two globin chains (the latter were not resolved from one another) were indeed the major translation products (Fig. 2, lane 1). Addition of 20 units of SRP caused inhibition of opsin synthesis without affecting the synthesis of globin (Fig. 2, lane 2). These data suggested that opsin contains a signal sequence which, by analogy to the cleaved NH₂-terminal signal sequences of secretory proteins¹⁹, caused an SRP-mediated elongation arrest. A discrete arrested fragment (or several discrete fragments if there are several signal sequences) has so far not been detected.

More definitive evidence for the existence of a signal sequence in opsin was the synthesis of a glycosylated form of opsin in the presence of SRP and KRM (SRP-depleted membranes) (Fig. 2, lane 4, downward-pointing arrow). As expected, this glycosylated form was sensitive to treatment with endoglycosidase H (Endo H) (Fig. 2, compare lanes 14 and 15). The synthesis of a glycosylated form of opsin indicated that at least the NH₂-terminal domain of opsin was translocated. Note that the glycosylated form of opsin was synthesized only when both KRM and SRP were present. In the absence of SRP only the non-glycosylated form was synthesized (Fig. 2, lane 3). The SRP-mediated inhibition of opsin synthesis was largely abolished in the presence of KRM (Fig. 2, compare lane 2 with lanes 3, 4). These data are consistent with release of an SRP-mediated elongation arrest by KRM¹⁹.

Although the SRP-dependent appearance of a glycosylated form of opsin suggested a signal sequence-mediated mechanism of translocation (at least for the glycosylated NH₂-terminal domain of opsin), it was conceivable that the three *trans* domains further downstream were integrated by insertion sequences (for example, independently of SRP). In this case, the non-glycosylated opsin synthesized in the presence of KRM only (Fig. 2, lane 3) might also be integrated. As a criterion for integration, we used non-extractability at alkaline pH (pH 11)^{8,20}. When synthesized in the absence of SRP, the non-glycosylated form of opsin was not integrated into KRM, as >90% was extracted by alkali (Fig. 2, compare lanes 5 and 6). The small amount of opsin that was resistant to alkali extraction was not integrated into KRM, but, rather, represented aggregated material co-sedimenting with KRM. This aggregated material was observed even when KRM and NaOH were added together (Fig. 2, lanes 9, 10) or when KRM were omitted altogether (Fig. 2, lanes 11, 12). Thus, it appears that insertion sequences are not used for the integration of opsin. However, these data alone do not exclude the possibility that translocation of the NH₂-terminal domain of opsin by a signal sequence is required for the subsequent expression of putative insertion sequences that may serve to integrate the other three translocated domains.

As expected, the glycosylated form of opsin synthesized in the presence of SRP was not extracted at alkaline pH (Fig. 2, compare lanes 7 and 8). In addition, when synthesized in the presence of SRP, a significant amount of the non-glycosylated form of opsin was resistant to alkali extraction. This non-glycosylated, non-extractable form of opsin probably represented opsin that had been integrated but not glycosylated. This could be a result of an inefficiency of dog pancreas microsomal membranes to glycosylate relative to their ability to translocate. Alternatively, if opsin possesses separate signal sequences for each translocated domain, the first one (responsible for translocation of the first *trans* (glycosylated) domain) might have been missed, yielding non-glycosylated opsin integrated via any or all of the other signal sequences. Note that all of the globin chains were completely extracted by alkali in any of the experimental conditions used (Fig. 2, lanes 5–12) and thus behaved as predicted for a non-integrated, cytoplasmic protein. Although resistance to alkali extraction does not prove that glycosylated opsin was integrated correctly (that is, with all four *trans* domains properly translocated), we can conclude that this was the case for at least the first *trans* domain, containing the two glycosylation sites.

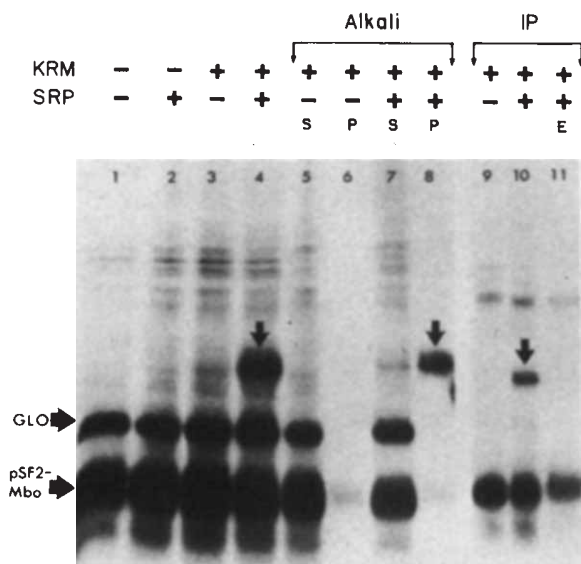


Fig. 3 The first 93 amino-terminal residues of opsin contain a signal sequence. pSF1 was digested with *Mbo*II and the resultant truncated plasmid (pSF2-Mbo) was used to generate mRNA translated in the wheat-germ system with (+) or without (-) SRP or KRM as described in Fig. 2 legend. The samples in lanes 5–8 were treated with alkali as described in Fig. 2 legend. The translation products in lanes 9–11 were immunoprecipitated with anti-opsin serum and those in lane 11 were treated with Endo H and then immunoprecipitated as described in Fig. 2 legend. Due to overexposure of this autoradiograph, the inhibition of synthesis of pSF2-Mbo is not readily apparent. When the globin and pSF2-Mbo bands were cut and quantitated¹⁹, the globin band in lane 1 contained 162,000 c.p.m. and that in lane 2 contained 171,000 c.p.m.; for pSF2-Mbo, these counts were 236,000 (lane 1) and 138,000 (lane 2). After correcting the pSF2-Mbo values for sample variability¹⁹, it was calculated that pSF2-Mbo synthesis was inhibited by 63% in the presence of SRP.

Localization of opsin signal sequence

Our initial strategy to localize the signal sequence was to determine the minimum length of opsin necessary for integration. The opsin-containing SP6 plasmid was cut with various restriction enzymes to generate a number of truncated opsins consisting of progressively shorter NH₂-terminal ends of the molecule. The shortest truncated form, pSF2-Mbo, encoded the 93 NH₂-terminal amino acids of opsin (calculated relative molecular mass (*M_r*) 12,300), and consisted of the first *trans* domain, the first transmembrane segment, the first *cis* domain and most of the second transmembrane segment. Translation yielded a product of the expected mass (Fig. 3, lane 1). Translation was inhibited in the presence of SRP (Fig. 3, compare lanes 1 and 2). Moreover, a glycosylated form of the 93-residue translation product was synthesized only in the presence of both KRM and SRP (Fig. 3, lane 4, downward-pointing arrow), not with KRM alone (Fig. 3, lane 3). As expected, the glycosylated form was not extracted by alkali (Fig. 3, compare lanes 7 and 8), whereas the non-glycosylated form synthesized in the absence of SRP was extracted (Fig. 3, compare lanes 5 and 6). The non-glycosylated and glycosylated forms of this polypeptide could be immunoprecipitated with an antiserum against a synthetic peptide representing residues 14–23 of opsin (Fig. 3, lanes 9–11). Furthermore, the glycosylated form was sensitive to Endo H digestion (Fig. 3, lane 11).

These data suggest that a signal sequence is localized in the first 93 residues of opsin. Of these, 30–40 carboxy-terminal residues can be expected to be sequestered in the ribosome^{21,22}. As the signal sequence is presumably expressed only when exposed on the surface of the ribosome¹⁹, the 30–40 carboxy-terminal residues of the 93-amino-acid polypeptide (which

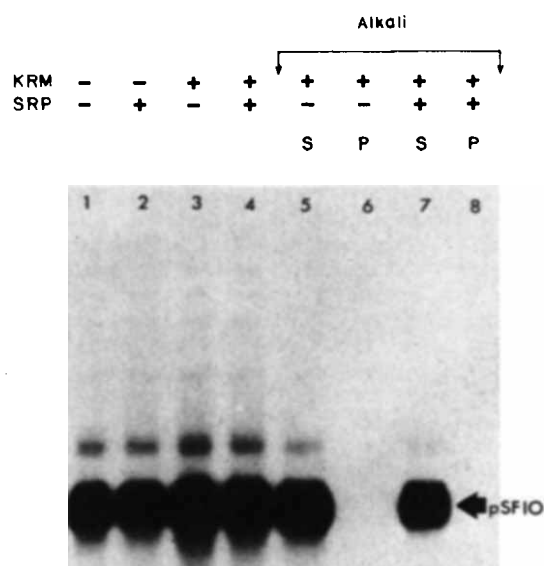


Fig. 4 The amino-terminal *trans* domain of opsin does not function as a signal sequence. A form of opsin containing only the amino-terminal 34 residues fused to the carboxy-terminal 37 residues (encoded by pSF10) was analysed for SRP-dependent integration and translocation. Reticulocyte RNA was omitted from the translation mixture because the proteins encoded by pSF10 and globin have similar electrophoretic mobilities (data not shown). pSF9-10 (300 ng) was used to programme the wheat-germ translation system in the presence (+) or absence (-) of KRM and SRP as described in Fig. 2 legend. The samples in lanes 5-8 were treated with alkali as described for Fig. 2.

include the second transmembrane segment and much of the first *cis* domain) could therefore not have served as a signal sequence. Thus, the first 53-63 residues (comprising the first *trans* domain and the first transmembrane segment) must contain the potential sites for a signal sequence.

To determine which of these two domains might contain a signal sequence, we constructed a plasmid (pSF10) in which most of the first *trans* domain was ligated to the fourth *cis* domain (the latter is highly charged and is not expected to contain a signal sequence). Since the translation product of this construct (calculated M_r 10,630) co-migrated with the globin chains (data not shown), reticulocyte RNA was omitted from the translation. The data shown in Fig. 4 indicate that this construct does not contain a signal sequence: there was no significant SRP-mediated inhibition of translation (Fig. 4, compare lanes 1 and 2), no SRP-mediated appearance of a glycosylated form (Fig. 4, compare lanes 3 and 4), and no SRP-mediated alkali-resistant integration (Fig. 4, compare lanes 5-8). We therefore conclude that a signal sequence is not localized in the region of the glycosylated *trans* domain, but, rather, in the region comprising the first transmembrane segment.

A second signal sequence

As discussed above, opsin could contain as many as four signal sequences, one for each domain to be translocated. To localize a second signal sequence, we ligated the DNA encoding the first *trans* domain to that encoding a carboxy-terminal portion of opsin beginning with the sixth transmembrane segment (Fig. 1). The transcribed RNA of this construct (pSF8) yielded a translation product with a predicted M_r of 16,500. It is clear that this opsin construct contains a signal sequence (Fig. 5a). Its translation was inhibited by SRP. Moreover, the synthesis of a glycosylated form (Fig. 5a, lane 4, downward-pointing arrow) that resisted alkali extraction was dependent on the presence of SRP (Fig. 5a, compare lanes 1-4). As neither the amino-terminal *trans* domain nor the carboxy-terminal *cis* domain contains a signal sequence (see Fig. 4), the signal

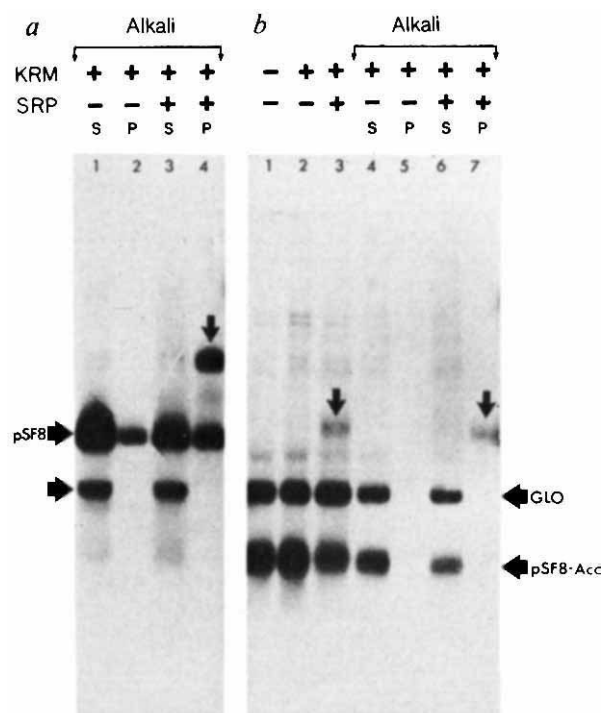


Fig. 5 A second signal sequence is located in the sixth transmembrane segment. *a*, An opsin construct (pSF8) containing the first and fourth *trans* domains, transmembrane segments VI and VII and the fourth *cis* domain, was analysed for SRP-dependent integration and translocation. pSF8 (150 ng) was used to programme the wheat-germ translation system in the presence (+) or absence (-) of KRM or SRP as described in Fig. 2 legend. Following translation, the samples were alkali-treated as described for Fig. 2. *b*, A truncated form of pSF8 (pSF8-Acc) (encoding the first *trans* domain, sixth transmembrane segment, *trans* loop 4 and 14 residues of the seventh transmembrane segment) was used to programme the wheat-germ translation system and the translation product was analysed for SRP-dependent integration and translocation as described for *a*. Following translation, the samples displayed in lanes 4-7 were extracted with alkali as described for Fig. 2. The glycosylated forms (downward-pointing arrows) were Endo H-sensitive (data not shown).

sequence of this construct must be localized in a region comprising transmembrane segments 6 and 7.

To localize the signal sequence more precisely, we prepared a truncated form of this construct (pSF8-Acc), resulting in the removal of the fourth *cis* domain and part of the seventh transmembrane segment. Of this translation product, 30-40 of the carboxy-terminal residues would not be exposed on the ribosome and therefore could not function as a signal sequence. Thus, if this construct contains a signal sequence, it must be localized in the sixth transmembrane segment, specifically in the amino-terminal portion of this segment. Figure 5b shows that a signal sequence was, in fact, present in this construct. A glycosylated form (Fig. 5b, lanes 3, 7, downward-pointing arrows) was synthesized only in the presence of SRP and was resistant to alkali extraction (Fig. 5b, compare lanes 4-7). We emphasize that although these data show the presence of a signal sequence in the amino-terminal portion of the sixth transmembrane segment, our experiments do not rule out the possibility that the seventh transmembrane segment also has the potential to function as a signal sequence.

In the two cases (Figs 3, 5b) where we have used truncated forms of opsin (lacking a termination codon), we have analysed the translation mixture by sucrose gradient centrifugation and subsequent SDS-polyacrylamide gel electrophoresis (PAGE) of gradient fractions. Surprisingly, in both cases ~60% of the truncated chains were released from the ribosomes (data not

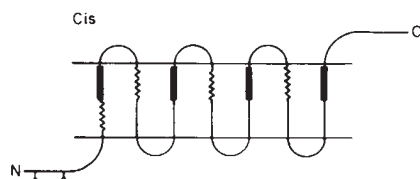


Fig. 6 Model illustrating the possible locations of signal and stop transfer sequences within opsin. See text for explanation. N, amino-terminus; C, carboxy-terminus; Cis, cytoplasmic side of membrane; zigzag lines, signal sequences; thickened portion of line, stop transfer sequences; λ , glycosylation sites.

shown). As a result, the 30–40 carboxy-terminal residues of these released chains were no longer located within the ribosome and, thus, a signal sequence located within this region could potentially interact with the membrane. However, the recent demonstration²³ that signal sequences addressed to the endoplasmic reticulum (ER) can be expressed only when presented to the ER in the context of the ribosome (that is, only when the chain is attached to the ribosome) rules out the possibility that the observed integration was due to the expression of a signal sequence exposed after chain release.

Perspectives

The most important conclusion from our data is that the integration of a polytopic integral membrane protein with multiple translocated domains may require multiple signal sequences. From hindsight, opsin turned out to be a fortunate choice. Its first *trans* domain of 36 residues contains the only two glycosylation sites (at residues 2 and 15) of the protein. We showed that this domain does not function as a signal sequence. In the deletion experiments described above, the first *trans* domain could always be retained and therefore be used as a convenient 'reporter' domain in the search for signal sequences elsewhere in the molecule; the readily detectable glycosylation of this domain in the *in vitro* translation-translocation system indicated that translocation of this domain had taken place. By definition, whenever this translocation event was SRP-dependent, a signal sequence mediated the event. So far we have localized two signal sequences, one in the first transmembrane segment, the other in the amino-terminal portion of the sixth transmembrane segment. We assume that in full-length opsin the first of these would function in the translocation of the first (glycosylated) *trans* domain and the latter in the translocation of the fourth *trans* domain. Signal sequences are also likely to be used for translocating the second and third *trans* domains and might be located in the amino-terminal portions of transmembrane segments 2 and 4 (Fig. 6).

Among the important unresolved questions is the precise topology of the various constructs in the microsomal membrane. It is likely, however, that the first NH₂-terminal *trans* domain, when glycosylated, is localized on the *trans* side of the membrane, since asparagine-linked glycosylation requires translocation²⁴.

Another unresolved question is how the signal sequence-induced translocation process (which goes to completion in the case of secretory proteins) is interrupted in the case of membrane proteins. One model proposes that a translocation is halted by a 'stop transfer' sequence which in turn triggers disassembly of a chain-conducting proteinaceous channel in the membrane¹¹. Alternative models propose that translocation is arrested by the interaction of hydrophobic stretches of the polypeptide chain with the lipid bilayer while the chain is passing across the bilayer (that is, the chain does not pass through a proteinaceous channel^{25,26}).

Figure 6 shows a scheme for the possible location of alternating signal and stop transfer sequences in opsin. The dynamics of the integration process could be envisaged within the framework of a previously proposed model¹¹. The interaction

of the first signal sequence with the membrane would proceed as described recently for secretory proteins (for summary, see ref. 4): the signal sequence, in the form of a loop^{26,27}, is recognized by SRP and a translation arrest is induced; the translation arrest is released by an interaction of the SRP with its cognate receptor (SRP receptor) in the ER, resulting in the displacement of SRP from the ribosome. These interactions bring the translating ribosome into the vicinity of the ER and would thus facilitate the subsequent interaction of the signal sequence with a putative signal receptor (which could be represented by signal peptidase) and of the ribosome with a putative ribosome receptor, both integral membrane proteins of the ER. The amino-terminal *trans* domain would then traverse the membrane through the proteinaceous channel (which might be formed, at least in part, by the signal receptor and the ribosome receptor). The signal sequence would be displaced from its ER receptor following further chain elongation. Passage of the chain would eventually be interrupted by the appearance in the channel of a stop transfer sequence. As a result, the channel would open to the lipid bilayer and the first integrated transmembrane segment would, after lateral displacement, be entirely surrounded by lipid molecules. Further chain elongation would result in the extra-ribosomal exposure of the second signal sequence (again in the form of a loop) which would initiate a second round of translocation. A second stop transfer sequence would stop the translocation again. Two more rounds of translocation-initiation and translocation-termination would result in the proper topology for opsin. It is conceivable that SRP and its cognate receptor in the ER are involved only in targeting the first signal sequence to the membrane. The second, third and fourth signal sequences, in a shortened circuit, may be able to interact directly with the signal receptor in the membrane without first interacting with SRP. Although it is unknown whether such a shortened circuit (which bypasses recognition of the signal sequence by SRP and uses only the putative signal receptor in the ER) could occur *in vivo*, integration *in vitro* mediated by the fourth signal sequence (even though tested here separately and out of context) requires SRP.

Many of the predictions illustrated in our model (Fig. 6) can be tested. Thus, the two additional predicted signal sequences could be localized by experiments similar to those described here. Similarly, it may be possible to separate the signal sequences from the putative stop transfer sequences by designing constructs that retain one of the signal sequences, but none of the stop transfer sequences. Such constructs would be expected to be completely translocated.

The approaches used here may serve as a paradigm for determining which topogenic sequences (insertion sequences, signal sequences) are used for the integration of polytopic integral membrane proteins and precisely where within the protein these sequences are located.

We thank Dr Jeremy Nathans for providing us with a bovine opsin cDNA clone, Stephen deSanctis for synthesizing the opsin peptides used for immunization, Drs James Kaput and Reid Gilmore for many helpful discussions, Drs Susan Bock, Ann Erickson, Sheila Fallon, Brian Kay and Debkumar Pain for helpful comments, Gisele Nimic for preparing the illustrations, and the Departments of Medicine and Medical Education at the Hackensack Medical Center. This work was supported by the Alexandrine and Alexander L. Sinsheimer Fund, NIH grant EY05540 to M.F. and NIH grant GM27155 to G.B.

Received 24 July; accepted 20 September 1985.

1. Katz, F. N., Rothman, J. E., Lingappa, V. R., Blobel, G. & Lodish, H. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3278–3282 (1977).
2. Toneguzzo, F. & Ghosh, H. P. *Proc. natn. Acad. Sci. U.S.A.* **75**, 715–719 (1978).
3. Lingappa, V. R., Katz, F. N., Lodish, H. F. & Blobel, G. *J. biol. Chem.* **253**, 8667–8670 (1978).
4. Walter, P., Gilmore, R. & Blobel, G. *Cell* **38**, 5–8 (1984).
5. Enoch, H. G., Fleming, P. J. & Strittmatter, P. *J. biol. Chem.* **254**, 6483–6488 (1979).
6. Tajima, S. & Sato, R. *J. Biochem., Tokyo* **87**, 123–134 (1980).
7. Takagaki, Y., Rhadakrishnan, R., Wirtz, K. W. A. & Khorana, H. G. *J. biol. Chem.* **258**, 9136–9142 (1983).
8. Anderson, D. J., Mostov, K. M. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7249–7253 (1983).

9. Sakaguchi, M., Mihara, K. & Sato, R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3361-3364 (1984).
10. Brown, D. A. & Simoni, R. D. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1674-1678 (1984).
11. Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1496-1500 (1980).
12. Hargrave, P. A. *et al. Biophys. struct. Mech.* **9**, 235-248 (1983).
13. Ovchinnikov, Y. A. *et al. Bioorg. Khim.* **8**, 1011-1014 (1982).
14. Nathans, J. & Hogness, D. *Cell* **34**, 807-814 (1983).
15. Schechter, I. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 2654-2658 (1979).
16. Goldman, B. M. & Blobel, G. *J. Cell Biol.* **90**, 236-242 (1981).
17. Melton, D. A. *et al. Nucleic Acids Res.* **12**, 7035-7056 (1984).
18. Erickson, A. & Blobel, G. *Meth. Enzym.* **96**, 38-49 (1983).
19. Walter, P. & Blobel, G. *J. Cell Biol.* **91**, 557-561 (1981).
20. Steck, T. L. & Yu, J. *J. supramolec. Struct.* **1**, 220-234 (1973).
21. Malkin, L. I. & Rich, A. *J. molec. Biol.* **26**, 329-346 (1967).
22. Blobel, G. & Sabatini, D. D. *J. Cell Biol.* **45**, 130-145 (1970).
23. Gilmore, R. & Blobel, G. *Cell* **42**, 497-505 (1985).
24. Snider, M. D. & Rogers, O. C. *Cell* **36**, 753-761 (1984).
25. Von Heijne, G. & Blomberg, C. *Eur. J. Biochem.* **97**, 175-181 (1979).
26. Engelman, D. M. & Steitz, T. A. *Cell* **23**, 411-422 (1981).
27. Inouye, M. & Halegoua, G. *CRC Crit. Rev. Biochem.* **10**, 339-371 (1980).
28. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning* (Cold Spring Harbor Laboratory, New York, 1982).
29. Walter, P. & Blobel, G. *Meth. Enzym.* **96**, 84-93 (1983).
30. Walter, P. & Blobel, G. *Meth. Enzym.* **96**, 682-691 (1983).
31. Anderson, D. J., Walter, P. & Blobel, G. *J. Cell Biol.* **93**, 501-506 (1982).

LETTERS TO NATURE

Radio emission from the jet and lobe of 3C273

R. J. Davis, T. W. B. Muxlow & R. G. Conway

Nuffield Radio Astronomy Laboratories, Jodrell Bank,
University of Manchester, Macclesfield, Cheshire SK11 9DL, UK

The 'superluminal' motion observed in the cores of radio sources such as 3C273 is now accepted^{1,2} as evidence of relativistic motion within a few parsecs of the centre, but it is less clear whether such speeds persist out to kiloparsec scales^{3,4}. The one-sidedness of such sources is often cited as evidence of relativistic Doppler beaming, but could equally be intrinsic. New maps of 3C273 at 151 and 408 MHz have been made using the MERLIN interferometer⁵, with dynamic range of $4 \times 10^3:1$ and $10^4:1$, respectively. These show that (1) there is an extended region or lobe to the south of the main jet, (2) the radio emission of the jet is continuous from the core to beyond the limit of the optical jet and (3) no counter-component can be found in the opposite direction to the jet. The ridge-line of the jet shows a 'wobble', the wavelength (λ) of which decreases by a factor of 6 along its length. We suggest a model in which the jet motion is relativistic throughout, the velocity decreasing from $\beta\gamma \approx 6$ to $\beta\gamma \approx 1$ (where β is the velocity and γ the Lorentz factor). Since the Doppler beaming factors calculated for this model are too small to hide the lobe or the head of a counter-component, we suggest that 3C273 is intrinsically one-sided.

The MERLIN interferometer⁵ was equipped for operation at 151 MHz in early 1985, and observations were made of 3C273 at this frequency on 1985 May 21. The resulting 151-MHz map at 3.0 arc s resolution is shown in Fig. 1 (the first MERLIN map to be published at this frequency), while Fig. 2 shows a 408-MHz map at 1.0 arc s resolution which was obtained by processing earlier observations⁶ using a new software procedure called MAP (R. G. Noble, A. Bailey and T.W.B.M., unpublished). This procedure, which uses the hybrid mapping method⁷, corrects both telescope errors and closure offsets in amplitude and phase. Furthermore, we have removed the effects of the confusing double source⁸, situated 8 arc min away from 3C273. These techniques result in a considerable improvement in dynamic range. A 408-MHz map was also made at 3 arc s resolution (not shown) for comparison with Fig. 1.

The previously known features of the radio source^{3,6,8} are the flat-spectrum core coincident with the quasar, which is unresolved at 1 arc s resolution, and the steep-spectrum radio jet extending 22 arc s in position angle 222.5°, leading to a bright head. (The head contains three sub-components⁹ which are merged at 1 arc s resolution.)

An entirely new feature in Figs 1 and 2 is an extended low-brightness region on the southern side of the jet. By analogy with classical double radio sources, we refer to this feature as the 'lobe'. There is no trace of extended emission on the northern side to the limit of measurement. In Fig. 2 the lobe has a well-defined outline about halfway along the jet, but there is additional extended emission close to the side of the jet out to

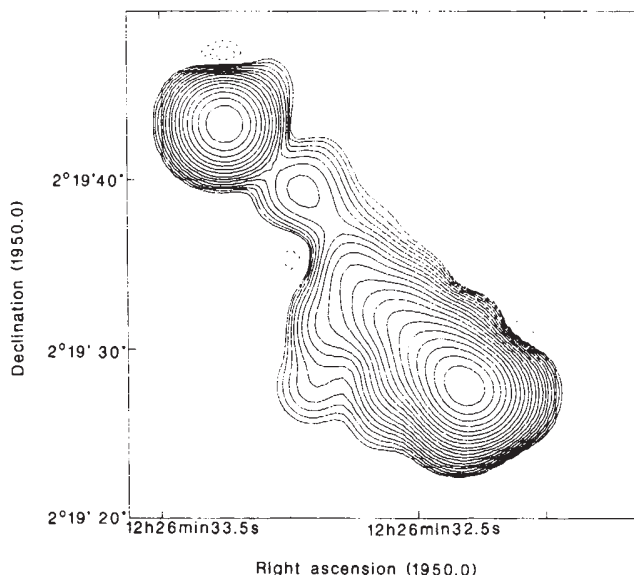


Fig. 1 MERLIN map of 3C273 at 151 MHz with 3.0 arc s resolution. Peak brightness is 55.8 Jy per beam. Contours are at equal logarithmic intervals, two contours to a factor 2, above a bottom contour of 56 mJy per beam. Negative (dashed) contours are at -56 and -80 mJy per beam.

20 arc s, appearing to spread backwards from the head at an angle of about 20°. At the lower resolution of Fig. 1, the 151-MHz lobe emission is blended with that of the main ridge, but is more intense and more extended than at 408 MHz. The spectral index α_{151}^{408} (obtained by comparing the two maps at 3 arc s resolution) is 1.5 ± 0.3 in the lobe and 0.80 ± 0.05 in the jet. From consideration of the integrated spectrum, Foley and Davis⁶ showed that the low-frequency radiation must come from a hitherto undetected area some 50 (arc s)² in area. This value matches the area of the outermost contour in Fig. 1, and we suggest that the lobe accounts for most of the flux at frequencies <20 MHz¹⁰.

The centre of the lobe is near the inner end of the optical jet, but there is no measured optical radiation from the lobe. (There is a blue extension on the north side¹¹, where there is no radio emission; this may be merely a coincidence.) The X-ray emission reported¹² as coming from the jet of 3C273 is also located near the inner end of the jet, and could arise in the radio lobe itself.

Both the morphology and the spectral index suggest that the lobe consists either of remnant material left behind after the passage of the active head, or of backflow. In either case, any motion will be slow, and the lobe emission will not be relativistically beamed. We have calculated physical parameters of the relativistic plasma in the lobe using standard formulae¹³, with $H_0 = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, so that 1 arc s = 1.85 kpc. We assume the line-of-sight thickness to be 5 kpc. If as much kinetic energy is assigned to the positive ions as to the electrons in the plasma, the minimum pressure in the lobe is $10^{-10} \text{ dyn cm}^{-2}$, the equipartition magnetic field is $\approx 60 \mu\text{G}$, and the synchrotron lifetime at 151 MHz is $5 \times 10^6 \text{ yr}$. The lobe must be at least this

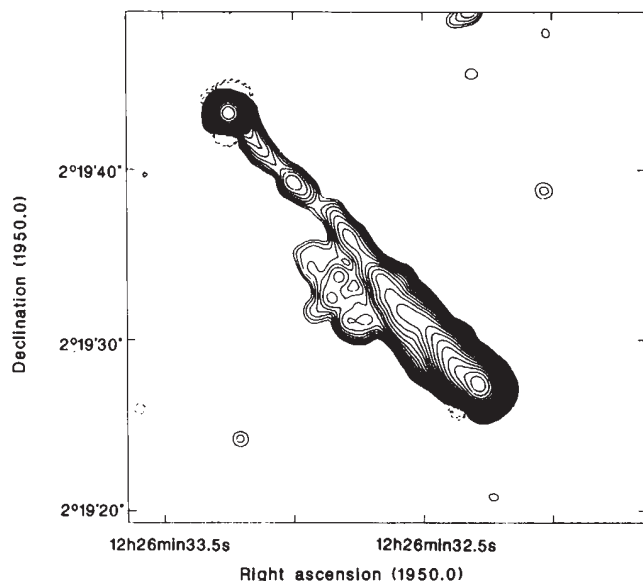


Fig. 2 MERLIN map of 3C273 at 407 MHz with 1.0 arc s resolution. Peak brightness is 13.5 Jy per beam. Contours are at equal logarithmic intervals, two contours to a factor 2, above a bottom contour of 6.75 mJy per beam. Negative (dashed) contours are at -6.7 and -9.5 mJy per beam. Positions in both Figs 1 and 2 have been calibrated by adjusting the position of the core to right ascension 12 h 26 min 33.248 s; declination +02° 19' 43.29" (VLA Calibrator list).

old if the steep spectral index indicates ageing. During this time, the head has moved forward some 10 arc s or 20 kpc, implying that the forward motion of the head is sub-relativistic, with a transverse component of velocity of the order of 0.01 c .

It might initially be supposed that the southward displacement of the lobe was caused by a transverse ambient 'wind' in the cluster of which 3C273 is a member¹⁴. However, the difficulty with this interpretation is that the ram pressure ρv^2 for the wind (where ρ is the ambient density and v the wind velocity) should be at least equal to the pressure in the lobe. If the wind velocity is of the order of 300 km s⁻¹, the required density corresponds to 10⁻¹ H atoms cm⁻³. Such a high density is hard to reconcile with the radio polarization data¹⁵: the rotation measure of all parts of the source is ≤ 1.5 rad m⁻², and hence the integral $\int n_e B dl$ in front of the source is $\leq 10^{-6}$ cm⁻³ G pc (where n_e is the electron density, B is the magnetic field and l is the path length). There is thus an unresolved discrepancy, unless the ambient intergalactic gas is singularly devoid of magnetic field ($B < 10^{-10}$ G over 100 kpc).

In all previous pictures, both radio and optical, the core and jet have been separated by a gap of 10 arc s. The new maps show that the radio emission of the jet continues across the gap as a faint narrow ridge, unresolved in width (angle $\theta < 1$ arc s), and with brightness $\approx 10^{-3}$ of the peak. In the brighter part of the jet, the width of the ridge apparently increases, but this may be solely due to blending with the lobe. The ridge can be traced over the whole 21 arc s of length, and Fig. 3a shows the intensity profile at 408 MHz along the ridge-line, plotted logarithmically. The rise to the peak follows an exponential increase with an e-folding scale height of ≈ 2 arc s, or 4 kpc. Similar profiles are found at 151 MHz (Fig. 1) and also at 1.67 GHz⁶ and 5 GHz⁸, with little or no change in scale height. If this region were a trail, subject to synchrotron ageing as in the lobe, the scale height would vary as $\nu^{0.5}$. We suggest instead that the intensity profile indicates the rate at which energy is extracted from bulk kinetic energy and dumped into the synchrotron reservoir.

In Fig. 3b the position of the ridge-line is plotted relative to an axis through the core at position angle 222.5°. Close to the head, the pattern can be followed best at 1.67 GHz, and the last five points are taken from the $\lambda 18$ cm map at 0.35 arc s resolution⁶. The 'wobble' in the ridge-line^{3,6} is present throughout the

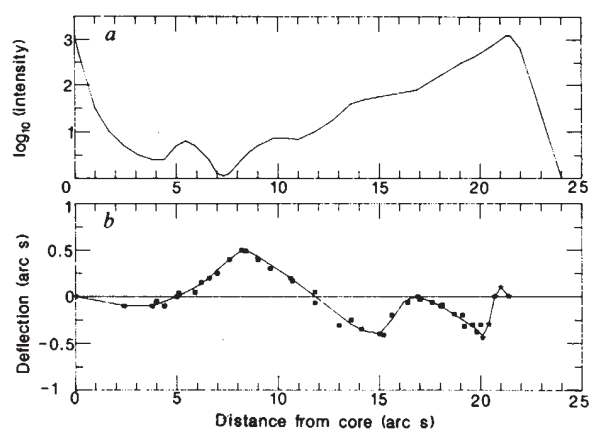


Fig. 3 a, Plot of intensity (log scale) along the ridge of the jet; b, deflection of ridge-line from a straight line through the core in position angle 222.5°. (The vertical scale is exaggerated $\times 5$.) ■, 408 MHz; *, 1,666 MHz.

length of the jet, superimposed on a centre-line which curves gently southwards by $\approx 2^\circ$ in position angle. The amplitude of the wiggle is constant, at about 0.5 arc s peak to peak, but the semi-wavelength of the pattern, measured between crossing-points, decreases systematically from 6 arc s to ≈ 1 arc s.

Wiggles are observed in many sources and have not been satisfactorily explained. However, the most straightforward interpretation of the systematic decrease in wavelength is as a deceleration of the jet material. If we interpret the wiggle in terms of precession in the central object with period T , the pattern wavelength λ is proportional to T and to the apparent transverse velocity β_{obs} :

$$\lambda = cT\beta_{\text{obs}} = cT\beta \sin \theta / (1 - \beta \cos \theta)$$

On this interpretation, the decrease of pattern wavelength down the jet indicates that deceleration begins at 10 arc s from the core, and continues to the head. The rate of deceleration thus mimics the rise in radio intensity. Because little braking is apparent in the inner part, we propose that the initial value of β_{obs} equals the superluminal value observed a few marc s from the core. The proper motion of 0.79 marc s yr⁻¹ (ref. 16) means that, if $H_0 = 100$ km s⁻¹ Mpc⁻¹, $\beta_{\text{obs}} = 4.8$ and T is 15×10^3 yr. Figure 3b then suggests that β_{obs} decreases to a value ≈ 0.8 at the point where the jet enters the head.

The degree of polarization observed at $\lambda 18$ cm in the brightest sub-component of the head⁹ has been explained in terms of relativistic aberration. If the Doppler blueshift factor is

$$\eta = (\gamma(1 - \beta \cos \theta))^{-1}$$

the aberration needed to account for the polarization (see ref. 9) requires that

$$\eta \sin \theta = \sin 47^\circ$$

By combining this value with the value of β_{obs} from the pattern wavelength, we may deduce the speed β and the angle θ of the motion near the head. If the value of θ does not change, parameters may also be obtained for the motion near the core.

Table 1 Model parameters for the motion in the jet reckoned in the frame of the observer

Region	Near core	Near head
Distance from core	≤ 10 arc s	21 arc s
Speed, β	0.987	0.74
Angle, θ	19°	19°
Lorentz, γ	6.2	1.5
Doppler blueshift, η	2.3	2.2
Apparent transverse velocity, β_{obs}	4.8	0.79
Factor $x = (1 + \beta \cos \theta) / (1 - \beta \cos \theta)$	28	5.5

Table 1 shows values for these quantities at each end of the jet, all values being expressed in the observer's frame of reference. The Lorentz factor of the initial velocity, when reckoned in the rest-frame of the quasar ($z = 0.158$), is $\gamma \approx 7$.

The velocity model presented here is consistent with the radio data, but is not unique. In particular, the parameters could vary widely in value if the jet were bent, that is, if θ varied along the length. We note that a bend of position angle occurs at 8 marc s from the core¹⁶, and that most of the measurements of superluminal motion refer to distances less than 8 marc s. A crucial test of our model would be provided by measurement beyond the bend. According to the model, the proper motion should remain at $0.79 \text{ marc s yr}^{-1}$ at all points out to 10 arc s, decreasing to $<0.1 \text{ marc s yr}^{-1}$ near the head. Further measurements of the proper motion at high dynamic range and long wavelengths would help to resolve this question.

The observations presented here may be used to test whether or not the one-sidedness of 3C273 is due to Doppler beaming. The observed peak brightness in the head at 408 MHz is 13 Jy per beam at 1 arc s resolution, and 22 Jy per beam at 3 arc s. Identifiable sidelobes due to residual errors are present at the 30-dB level (1,000:1 down) north and south of bright features. Away from these points the r.m.s. noise level is 1.2 mJy per beam at 1 arc s resolution, or 1.9 mJy per beam at 3 arc s (these values are the observed r.m.s. values over several hundred square arc s in empty regions of the map, and are $\approx 50\%$ above the calculated thermal noise). Thus, the observed dynamic range at 408 MHz (defined as peak intensity per r.m.s. level) is $10^4:1$. At 151 MHz the corresponding values are 56 Jy per beam peak brightness, r.m.s. level 13 mJy per beam, giving a dynamic range of $4 \times 10^3:1$. Along the continuation of the major axis from the quasar out to the geometrical counter-head position 22 arc s away the 408-MHz map shows no emission above noise level, with a maximum value of 4 mJy per beam (2σ) near the geometrical counter-head.

We have calculated the intrinsic front/back ratio in luminosity for each component of the source, allowing for the Doppler beaming factors of any that are fast-moving. At the head the observed brightness at 3 arc s resolution is 22 Jy per beam, so that the observed head/counter-head ratio is $>5,500:1$. The expected brightness ratio for two equal jets in opposite directions is $x^{2+\alpha}$, where $x = (1 + \beta \cos \theta)/(1 - \beta \cos \theta)$. For the model suggested above with $\beta = 0.74$ and $\theta = 19^\circ$, the Doppler beaming ratio is 120. Even if the entire radiation from the head were beamed with this Doppler factor, the expected intensity from a counter-jet would be 190 mJy per beam, making the front/back ratio for the jet $\geq 47:1$. In practice the head almost certainly emits a mixture of beamed and unbeamed radiation. Flatters and Conway⁹ suggested that 1/6th of the head flux comes from a slow-moving precursor component with little beaming. The intrinsic front/back ratio on this basis is $\geq 1,000:1$.

The lobe emits at least 500 mJy unbeamed radiation at 408 MHz from an area of 15 (arc s)², while no region of equal area on the opposite side of the core emits more than 10 mJy. Thus, the observed front/back ratio for the lobe is $\geq 50:1$, in sharp distinction from classical double sources. While the high front/back ratios for the head are model-dependent, being based on our assumptions of jet speed, the front/back ratio for the lobe is free from such assumptions.

We conclude that there is no observational evidence of any counter-components in 3C273. Doppler beaming is insufficient to hide the counterpart of the lobe and probably that of the head. If such counterparts do exist, they must be intrinsically weaker than those observed by one or two orders of magnitude. Thus, 3C273 contains all the ingredients of a classical double source like Cygnus A, with core, jet, hotspots and extended lobe, but only on one side.

We thank R. A. Perley for data in advance of publication, and C. Flatters for useful discussions. The new high-quality maps were only made possible by software developed by R. G. Noble and A. Bailey on the new Starlink VAX node at Jodrell Bank.

Received 22 July; accepted 4 October 1985.

1. Blandford, R. D. & Konigl, A. *Astrophys. J.* **232**, 34-48 (1979).
2. Begelman, M. C., Blandford, R. D. & Rees, M. J. *Rev. mod. Phys.* **56**, No. 2 Pt 1, 255-351 (1984).
3. Conway, R. G., Davis, R. J., Foley, A. R. & Ray, T. P. *Nature* **294**, 540-542 (1981).
4. Bridle, A. H. & Perley, R. A. *Rev. Astr. Astrophys.* **22**, 319-358 (1984).
5. Davies, J. G., Anderson, B. & Morison, I. *Nature* **288**, 64-66 (1980).
6. Foley, A. R. & Davis, R. J. *Mon. Not. R. astr. Soc.* **216**, 679-690 (1985).
7. Cornwell, T. J. & Wilkinson, P. N. *Mon. Not. R. astr. Soc.* **196**, 1067-1086 (1981).
8. Perley, R. A. *Proc. IAU Symp.* **110**, 153-156 (1984).
9. Flatters, C. & Conway, R. G. *Nature* **314**, 425-426 (1985).
10. Braude, S. Y., Megn, A. V., Sokolov, K. P., Tkachenko, A. P. & Sharykin, N. K. *Astrophys. Space Sci.* **64**, 73-126 (1979).
11. Lelievre, G., Nieto, J.-L., Horville, D., Renard, L. & Servan, B. *Astr. Astrophys.* **138**, 49-56 (1984).
12. Willingale, R. *Mon. Not. R. astr. Soc.* **194**, 359-364 (1981).
13. Conway, R. G. *et al. Mon. Not. R. astr. Soc.* **202**, 813-823 (1983).
14. Tyson, J. A., Baum, W. A. & Kreidl, T. *Astrophys. J.* **257**, L1-L5 (1982).
15. Conway, R. G. & Stannard, D. *Nature phys. Sci.* **239**, 22-23 (1972).
16. Unwin, S. C. *et al. Astrophys. J.* **289**, 109-119 (1985).

Implications of ultra-high-energy emission from Hercules X-1

David Eichler

Astronomy Program, University of Maryland, College Park, Maryland 20742, USA and Department of Physics, Ben Gurion University, Beer Sheva, Israel

W. Thomas Vestrand

Physics Department, University of New Hampshire, Durham, New Hampshire 03824, USA

A 3-min outburst of very-high energy (VHE) quanta, $E > 10^{12} \text{ eV}$, was recently detected from the X-ray pulsar Hercules X-1 (ref. 1). The outburst occurred at a time during the 35-day X-ray modulation that is associated with X-ray turn-on. Temporal analysis of the outburst showed the same 1.24-s modulation that is observed at X-ray energies. Subsequent monitoring of the system at ultra-high energies (UHE) by the Fly's Eye² yielded evidence for a 40-min outburst of quanta with energies $E > 1 \times 10^{14} \text{ TeV}$. This UHE outburst also exhibited a 1.24-s modulation. Here we assume the quanta to be γ -rays and show how the interaction of these UHE particles with a precessing accretion disk can explain the observed γ -ray 'light' curve. We also discuss the possibility that the emitting UHE particles are accelerated by shocks in an accretion flow as well as the constraints that can be placed on the acceleration mechanism.

Hercules X-1 displays X-ray modulation with periods of 1.24 s, ~ 1.7 days and ~ 35 days³⁻⁵. The simplest interpretation of the two shorter periodicities is that they are due to rotation and occultation of an accreting neutron star located in a close binary system³⁻⁷. The unusual 35-day flux modulation has an X-ray light curve that is composed of an 11-day high-intensity state and a 19-day low-intensity state, interrupted midway between the 11-day high states by an intermediate high state (intensity $\sim 40\%$ of main high state) of 5 days duration^{3,4}. This modulation is usually referred to by the varying aspect of an inclined accretion disk that precesses about the X-ray pulsar⁵⁻⁷. In this picture, the 19-day low state occurs when our view of the pulsar is obscured by the accretion disk and the 11-day high state occurs when our view is unobstructed. The intermediate high state occurs when our line-of-sight passes close to the disk and is partially obscured by the disk's corona.

An intriguing aspect of the high-energy observations is the phase at which the γ -ray outbursts occur during the 35-day X-ray modulation. The UHE outburst occurred on 11 July 1983 at phase $\phi_{35} = 0.63$ in the 35-day cycle². This epoch corresponds to the middle of the period, when the intermediate high state is normally observed. The VHE outburst occurred on 17 April 1983¹, the nominal time for the onset of the 11-day high-intensity state. Both outbursts occurred at a time when the accretion disk is situated so that our line-of-sight was grazing the disk.

An attractive explanation for these observations is provided by a particular variation in the class of beam dump models⁸ in which the accretion disk serves as the beam dump. It is related to the model we later proposed⁹⁻¹¹ for the reported post-eclipse emission of Cygnus X-3, which has also been applied to LMCX-4 (ref. 12). The model for Her X-1 assumes that UHE particles are accelerated in a co-rotating region near the pulsar and then stream outward to interact with the surrounding accretion disk. Energetic γ -rays will then be detectable when the beam of particles crosses our line-of-sight and interacts with a column thickness of material that is comparable to the particle's radiation length. A smaller thickness would be inefficient as a converter, whereas a larger column thickness would obscure the photons. In the Her X-1 system, this condition is usually best met at the onset and decline of the X-ray high states when our line-of-sight is grazing the precessing accretion disk. In fact, the VHE outburst did occur at the nominal time for the onset of the 11-day X-ray high state. However, the production of the UHE outburst during a phase normally associated with the centre of the secondary X-ray high state would require a thickening of the disk to yield sufficient target along our line-of-sight. The Exosat X-ray satellite observations of the source during the same period support this idea¹³. The satellite's failure to detect the X-ray high states, despite the continuing optical variability attributed to the reprocessing of X rays, led several authors^{13,14} to conclude independently that the disk was significantly thicker during this epoch.

The duty cycle of the UHE emission is only about 10% of the spin period², which we take to imply beaming of the UHE primaries.

We next consider the efficiency of particle acceleration in the Her X-1 system. The median energy of 500 TeV quoted by the Fly's Eye group² is based on the assumption that the γ rays have a power law spectrum. The apparatus in fact responded to γ rays $> 1 \times 10^{14}$ eV (Elbert, J. W., personal communication). Taking, as a conservative estimate, the photon energies to be $E = 1 \times 10^{14}$ eV, a reported time-averaged UHE photon flux of $3 \times 10^{-12} \text{ cm}^{-2} \text{ s}^{-1}$ (ref. 2) and a duty cycle of 0.1, we derive a peak UHE energy flux of $3 \times 10^3 \text{ eV cm}^{-2}$. This should be compared with the total X-ray flux of $4 \times 10^3 \text{ eV cm}^{-2}$. The conversion efficiency for UHE particles to photons is only $\sim 10\%$. A liberal estimate for the beaming factor can be made by assuming that the broadest part of the beam passes through our line-of-sight. The solid angle of the beam is then $\pi^3 \times 10^{-2}$, and the beaming factor is $\pi^2/4 \times 10^{-2}$. We conclude that the UHE particle luminosity is at least $\sim 25\%$ of the X-ray luminosity.

This sets strong constraints on models of particle acceleration. For example, acceleration scenarios¹⁵ that invoke the rotational energy of the accretion disk encounter the difficulty of that for Her X-1 parameters, which include a surface magnetic field of 4×10^{12} G, the inner radius of the accretion disk is located at more than 300 stellar radii. It is difficult to see how more than about 1/300th of the total energy budget could be channelled through rotation of the accretion disk. The fact that the UHE and VHE emission has the periodicity of the neutron star's rotation further supports the notion that the energy that goes into the UHE particles is liberated close enough to the neutron star for the in-falling material to co-rotate with it.

The high efficiency associated with shock acceleration makes it an attractive possibility for particle acceleration in the Her X-1 system. In particular, one expects a shock from the impact of the accreting material onto the star or its magnetosphere. Given that spontaneous injection occurs¹⁶, shock acceleration can take place¹⁷. It is not necessary for the shock to occur at the surface of the neutron star for the mechanism to put much of the released gravitational energy into UHE particles; all that is needed is that much of the pressure in the post-shock material that settles on the surface is in the form of UHE particles trapped in the flow. However, our understanding of shock acceleration enables us to place strong constraints on models that invoke it.

The synchrotron loss timescale for a particle of mass m , energy

$E_p = \gamma mc^2$ and charge Ze is given by

$$\tau_{\text{sy}}(E_p) = \frac{4\pi mc}{\gamma \sigma_T B^2} \left(\frac{m}{Z^2 m_e} \right)^2 \quad (1)$$

where B is the magnetic field strength in the region and σ_T is the Thompson cross-section. The time required to accelerate a relativistic particle to energy E_p is

$$\tau_a(E_p) = (\xi R_g) / (\beta_s^2 c) \quad (2)$$

where R_g is the gyroradius ($R_g = \gamma mc^2 / (ZeB)$), ξR_g is the mean free path ($\xi \geq 1$) and $\beta_s c$ is the shock velocity. The maximum energy E_d to which a particle can be accelerated by a shock, even in the absence of synchrotron losses, is given by¹⁸

$$E_d = \frac{1}{\pi} \beta_s ZeBR \quad (3)$$

where R is the radius of the shock, which must be less than the size of the region. This limit comes from the fact that particles with energies greater than E_d can diffuse away from the system within the acceleration timescale. Stipulating that the acceleration must be faster than the synchrotron loss, and combining this limit with equation (3), we find that the maximum energy $\gamma_m mc^2$ to which a particle can be accelerated within a compact region of size R is given by

$$\gamma_m = \beta_s \left[\left(\frac{3}{2\pi\xi} \right) \frac{Rm}{Z^2 r_0 m_e} \right]^{1/3} \quad (4)$$

where r_0 is the classical electron radius.

If the shock velocity is taken to be the free fall velocity at a radius R from the neutron star, $V_{\text{ff}} = (2GM_*/R)^{1/2}$, then individual particle energies as high as $9 \times 10^6 mc^2 R_6^{-1/6}$ are possible (where $R = R_6 \times 10^{-6} \text{ cm}$). This is sufficient to account for the UHE emission from Her X-1. Note also that this upper limit is rather pessimistic because it uses a classical equation for the synchrotron loss timescale. In fact, this loss rate is an overestimate when radiation reaction and quantum mechanical effects become important.

A final constraint, we note, is that the Alfvén Mach number of the shock must be high if much of the energy is to go into the highest-energy particles. That is

$$\rho_s u_s^2 \gg \frac{B^2}{8\pi} \quad (5)$$

where ρ_s is the preshock fluid density and u_s is $\beta_s c$. If ρ_s is fixed by the condition that

$$\rho_s u_s (\pi R^2) = M = \frac{L}{(\epsilon c^2)} \quad (6)$$

where M is the accretion rate, and is the conversion efficiency from accreted mass to luminosity L , this constrains the magnetic field to be less than $\beta_s^{1/2} L_{38}^{1/2} R_6^{-1} \times 10^8 \text{ G}$, and R to be less than 10^8 cm . The synchrotron loss limit also constrains B to be less than $7 \times 10^{21} \xi \gamma_m^{-2} \text{ G}$. We see then that the model demands that the magnetic field within the accretion column be much less than the surface field of the neutron star. This is reasonable because the accreting material is a good conductor and is likely to pass between field lines of the neutron star on its way to the surface. However, it is not clear to what extent the flow retains its coherence between the magnetopause and the stellar surface. Alternatively, one could invoke a shock at 10–50 neutron star radii, where the field strength of the star's magnetosphere is 10^7 – 10^9 G .

We have proposed that the UHE emission reported from Her X-1 is generated by particles that interact with the surrounding accretion disk. The model predicts that high-energy γ -ray outbursts should occur preferentially at the onset and decline of the high-intensity X-ray states. The acceleration of the primaries occurs within the neutron star's magnetosphere, and the most obvious mechanism is acceleration by the accretion shock. A substantial fraction of the γ -rays may be generated within the

accretion column, the relative contribution of these two targets depending on the γ - γ opacity. Accretion-shock acceleration can account for the highest observed γ -ray energies, but not by a wide margin, and would be ruled out by detection of photons at 10^{16} eV or higher. Photons at $E > 10^{16}$ eV have been claimed by several authors, implying primary energies of $\sim 10^{17}$ eV (refs 12, 19). If these energy measurements are correct to within a factor of two, and if deep inelastic scattering behaves at these high energies as at lower ones, then according to our analysis, acceleration by accretion shocks cannot account for the primary energies. We emphasize the importance of precise energy measurements; γ -ray energies of $2\text{--}3 \times 10^{15}$ eV are marginally consistent with accretion-shock acceleration of the primaries. On the other hand, a shock created by a relativistic wind striking, say, an accretion disk at $R > 10^8$ cm, could generate primary energies of 10^{17} eV. Relativistic winds could result from supercritical accretion. It might be useful therefore, to search for correlations between E_{max} and evidence for relativistic winds.

Received 19 July; accepted 19 September 1985.

1. Dowthwaite, J. *et al.* *Nature* **309**, 691–693 (1984).
2. Baltrusaitis, R. M. *et al.* Preprint, Univ. Utah (1985).
3. Giacconi, R. *et al.* *Astrophys. J.* **184**, 227–236 (1973).
4. Jones, C. & Forman, W. *Astrophys. J. Lett.* **209**, L131–L135 (1976).
5. Petterson, J. A. *Astrophys. J. Lett.* **201**, L61–L64 (1975).
6. Petterson, J. A. *Astrophys. J.* **218**, 783–791 (1977).
7. Boynton, P., Crosa, L. & Deeter, J. *Astrophys. J.* **237**, 169–174 (1980).
8. Eichler, D. *Nature* **275**, 725–726 (1978).
9. Vestrand, W. T. & Eichler, D. *Proc. of the Workshop on Particle Acceleration in Astrophysics, La Jolla* (eds Arons, M. & McKee) 285–288 (1979).
10. Eichler, D. & Vestrand, W. T. *Proc. of the 16th International Cosmic Ray Conf., Kyoto*, **1**, 147–149 (1979).
11. Vestrand, W. T. & Eichler, D. *Astrophys. J.* **261**, 251–258 (1982).
12. Protheroe, R. J. & Clay, R. W. *Nature* **315**, 205–207 (1985).
13. Parmar, A. N. *et al.* *Nature*, **313**, 119–121 (1985).
14. Delgado, A. J., Schmidt, H. U. & Thomas, H. C. *Astr. Astrophys. Lett.* **127**, L15–L16 (1983).
15. Channugam, G. & Brecher, K. *Nature* **313**, 767–768 (1985).
16. Eichler, D. *Astrophys. J.* **229**, 419 (1979).
17. Cowsik, R. & Lee, M. A. *Proc. R. Soc. A* **383**, 409–423 (1982).
18. Eichler, D. *Astrophys. J.* **244**, 711–716 (1981).
19. Samorski, M. & Stamm, W. *Astrophys. J.* **268**, L17–L21 (1985).

Relationship between peroxyacetyl nitrate and nitrogen oxides in the clean troposphere

H. B. Singh*, L. J. Salas*, B. A. Ridley†, J. D. Shetter†, N. M. Donahue†, F. C. Fehsenfeld‡§, D. W. Fahey‡, D. D. Parrish§, E. J. Williams§, S. C. Liu‡, G. Hübler§ & P. C. Murphy§

* SRI International, Menlo Park, California 94025, USA

† National Center for Atmospheric Research, Boulder, Colorado 80307, USA

‡ National Oceanic and Atmospheric Administration, Aeronomy Laboratory, Boulder, Colorado 80303, USA

§ Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado 80309, USA

Nitrogen oxides have a key role in the chemistry of the polluted and the unpolluted atmosphere^{1–4}. It has been proposed that important quantities of nitrogen oxides may be contained in the form of peroxyacetyl nitrate (PAN) especially in the colder regions of the middle and upper troposphere^{5,6}. Surface and aircraft observations of PAN have confirmed its ubiquity in the remote troposphere where it has been measured at mixing ratios of 10–500 p.p.t.v. (parts per 10^{12} v/v) (refs 7, 8). To date, no study that is representative of the remote atmosphere has been made to assess directly the concentrations of PAN compared with that of the inorganic nitrogen oxides (NO and NO₂). Here we present the first study in which the mixing ratios of PAN, and nitrogen oxides—as well

as those of peroxypropionyl nitrate (PPN) and O₃ and relevant meteorological parameters—were measured concurrently at a location that receives clean, continental air. The results show that in clean conditions nitrogen oxides present in the form of PAN can be as much or more abundant than the inorganic form.

The measurements were made at the C-1 research site of the Mountain Research Station of the University of Colorado (40°2' N lat., 105°32' W long., 3.05-km elevation above sea level). Prevailing winds at this site are from the west and are relatively free of anthropogenic pollution. However, there are frequent occasions when winds from the east transport anthropogenic pollution to the site. The measurements, then, are representative of very clean to polluted conditions and must be interpreted in the context of the prevailing meteorology. This field study was conducted during the summer (1–16 July 1984; 24 August–2 September 1984) and the autumn (24 October–4 November 1984). The autumn period resembled early winter in that snow cover, high winds and low temperatures were typical.

The instruments used to measure NO_x, PAN and PPN, and O₃ have been described elsewhere^{9–11}. Briefly, NO was measured with a chemiluminescent technique and NO₂ was photolysed to NO before detection. The instrument has detection limits of 2 p.p.t. for NO and 10 p.p.t. for NO₂ based on a 10-s averaging time. The absolute accuracy of these measurements is estimated to be 15% for NO, 30% for NO₂ and 40% for NO_x. Unlike many instruments that catalytically reduce NO₂ to NO, the interference of PAN in the present NO₂ photolysis system was found to be <5% in laboratory tests^{9,10}. PAN and PPN were measured with two different instruments, both employing electron capture gas chromatography and cryogenic sample collection¹. Although the two PAN instruments did not operate concurrently, an intercomparison of results during similar atmospheric conditions indicated no systematic differences. With a 100 cm³ (STP) sample the PAN, PPN instruments had a detection limit of ~ 5 p.p.t., a precision of $\pm 10\%$ and an estimated accuracy of 20–30%. O₃ was measured by a UV-absorption instrument (Dasibi, Inc., Model 1003-AH) with a detection limit of 3 p.p.b.v. (parts per 10^9 v/v) and an estimated uncertainty of $\pm 10\%$.

The relationship between the mixing ratios of PAN and NO_x are examined in Fig. 1. Summer and autumn data from two periods of the diurnal cycle are presented. First is an afternoon period (1100 to 1800 MST) that is characteristic of the results of the day's photochemical activity. Second is the photochemically- and meteorologically-quietest night-time period (000 to 800 MST) of late night to early morning. In the summer afternoon, [PAN] increases with increasing [NO_x]. This behaviour is expected as PAN is rapidly produced by photochemistry in summer, at least in polluted air masses, and [NO_x] is a good indicator of the degree of urban pollution at this site⁹. In contrast, in the autumn and in the summer night-time the abundance of PAN is nearly constant and independent of the NO_x mixing ratio.

Figure 2 shows the mean diurnal behaviour of [PAN]. The symbols show the average PAN mixing ratio for periods when the NO_x mixing ratio was 300 p.p.t.v. or less. This [NO_x] criterion generally coincides with westerly winds at the site and is chosen as representative of relatively clean continental air masses. Figure 2 indicates that relatively constant PAN levels are encountered at this site in the summer and in the autumn in clean atmospheric conditions. The absence of a distinct diurnal cycle in the PAN levels indicates that PAN production and loss are approximately in balance throughout the day in clean continental air in the boundary layer troposphere. For these clean air masses the daily average PAN level was higher in the autumn (243 p.p.t.v.) than in the summer (170 p.p.t.v.). The PPN data gave results similar to those for PAN during both seasons at all NO_x levels with an average [PPN]/[PAN] ratio of 0.05.

For comparison, the mean diurnal behaviour of all the [PAN] data for these two seasons is also presented in Fig. 2 and is shown as the dashed lines. Here, the PAN mixing ratio does show a strong afternoon maximum in the summer but not in

|| Present addresses: NASA Ames Research Center, Moffett Field, California 94035, USA (H.B.S. and L.J.S.).

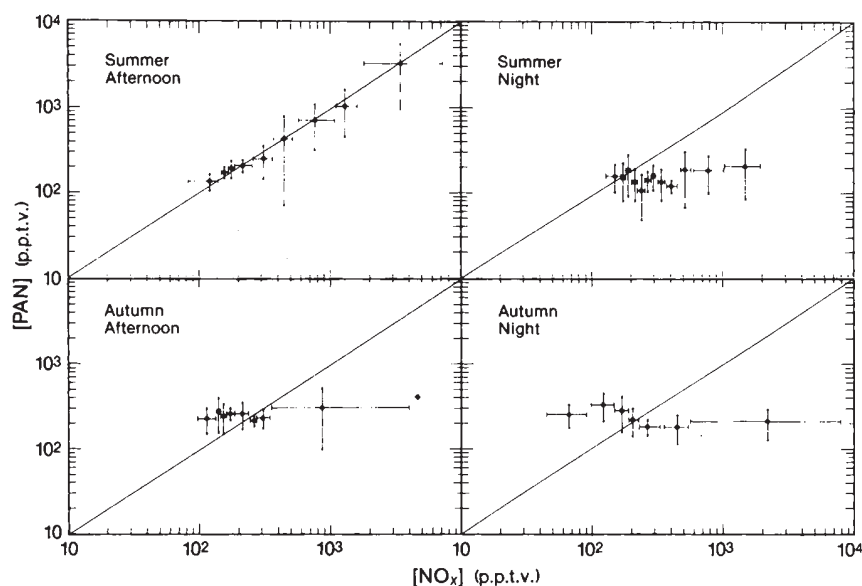


Fig. 1 PAN mixing ratio as a function of the NO_x mixing ratio. For both the summer and autumn sampling periods, afternoon (1100–1800 MST) and night-time (000–800 MST) data are given. Each set of data was arranged in order of increasing $[\text{NO}_x]$. Each symbol gives the $[\text{PAN}]$ and $[\text{NO}_x]$ averages of 10 consecutive $[\text{NO}_x]$ values. The vertical bars indicate the standard deviation ($\pm 1\sigma$) of the $[\text{PAN}]$ values and the horizontal bars show the range of NO_x mixing ratios included in each average. The diagonal lines indicate equal mixing ratios for PAN and NO_x .

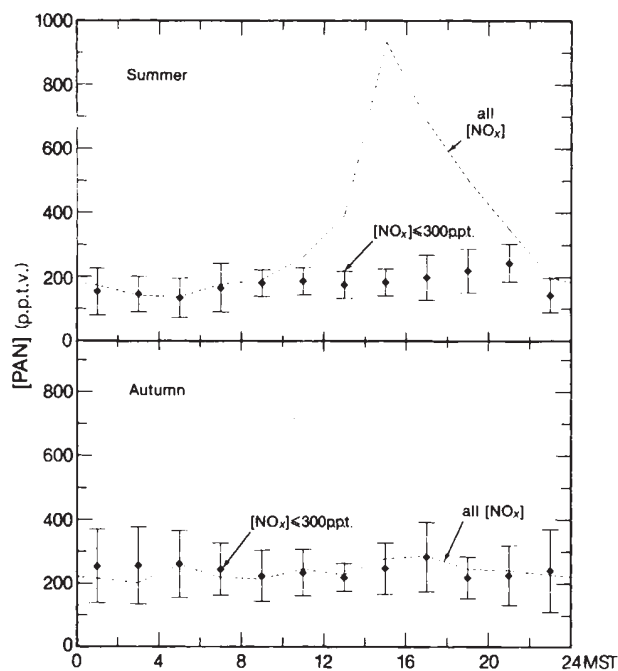


Fig. 2 Diurnal variation of PAN mixing ratio in the summer and in the autumn. The symbols represent averages of $[\text{PAN}]$ for a 2-h period of the diurnal cycle for the subset of the total measurements that is defined by $[\text{NO}_x] \leq 300$ p.p.t.v. The vertical bars indicate the standard deviation ($\pm 1\sigma$) of the $[\text{PAN}]$ values. The dashed lines given the corresponding averages for the full data set (all $[\text{NO}_x]$ values).

the autumn. This summer peak is attributed to both photochemistry and meteorology. During the summer, along the Front Range in the Colorado Mountains, the easterly, upslope winds increase in frequency during the daytime as a result of the daytime heating of the mountains. These easterly winds bring air containing recent (≤ 12 h) additions of pollution from the Denver metropolitan area, including high concentrations of both PAN and NO_x . The summer afternoon PAN peak observed at this site is similar in shape to that observed by Lewis *et al.*¹² in the summer in New Jersey, but with lower absolute values.

The importance of PAN relative to NO_x is shown in Fig. 3. The $[\text{PAN}]$ to $[\text{NO}_x]$ ratio increases with decreasing $[\text{NO}_x]$ values below 500 p.p.t.v. in both summer and autumn. This increase, particularly in the autumn, is principally due to the

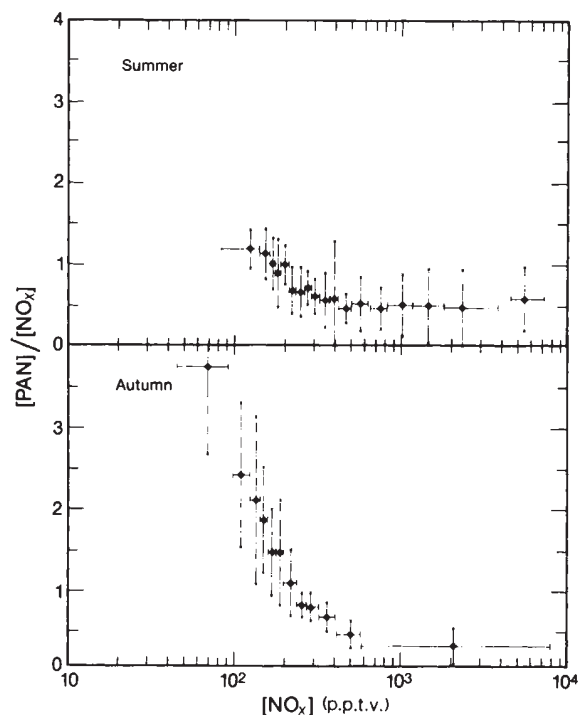


Fig. 3 $[\text{PAN}]/[\text{NO}_x]$ ratio as a function of $[\text{NO}_x]$ for the summer and the autumn. Each data set was arranged in order of increasing $[\text{NO}_x]$. Each symbol gives the $[\text{PAN}]/[\text{NO}_x]$ and $[\text{NO}_x]$ averages of 20 consecutive $[\text{NO}_x]$ values. The vertical bars indicate the standard deviation ($\pm 1\sigma$) of the ratio values while the horizontal bars show the range of $[\text{NO}_x]$ included in each average.

PAN levels remaining constant as $[\text{NO}_x]$ decreases. For NO_x levels < 300 p.p.t.v., the average $[\text{PAN}]/[\text{NO}_x]$ ratios are 0.5 to ≥ 1 in summer and 0.5 to ≥ 3 in autumn. The higher $[\text{PAN}]/[\text{NO}_x]$ ratio in autumn can be attributed, at least in part, to the increased stability of PAN at the lower autumn temperatures. No other determinations of the $[\text{PAN}]/[\text{NO}_x]$ ratio at comparable NO_x mixing ratios have been reported. However, there are other determinations at higher NO_x levels^{13,14} which are consistent with our results.

Figure 4 shows the observed relationship between the PAN and O_3 mixing ratios. High PAN levels are associated with high O_3 concentrations in the summer but the two are essentially decoupled in the autumn. This relationship between O_3 and PAN suggests that both PAN and O_3 are produced photochemi-

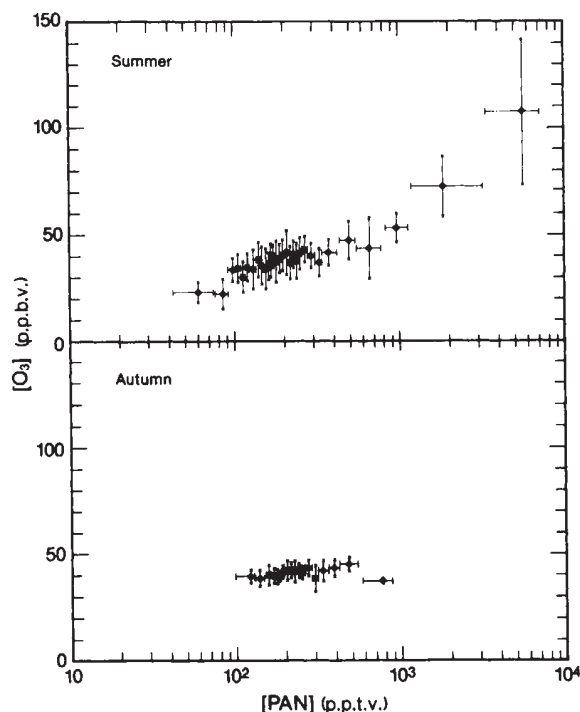


Fig. 4 Ozone mixing ratio as a function of [PAN] for the summer and the autumn. Each data set was arranged in order of increasing [PAN]. Each symbol then gives the [O₃] and [PAN] averages of 15 consecutive [PAN] values. The vertical bars indicate the standard deviation ($\pm 1\sigma$) of the [O₃] values while the horizontal bars show the range of [PAN] included in each average.

cally near the site. The photochemical production of O₃ at this site has been observed and reported earlier^{15,16}. The independence of the PAN, O₃ and NO_x concentrations in the autumn is attributed to slower photochemistry.

The precursors of PAN are peroxyacetyl (CH₃C(O)OO, or PA) radicals and NO₂. The PA radicals are produced by the reaction of acetaldehyde and hydroxyl radicals, by the photolysis of acetone and by the reactions of methylvinyl ketone and methacrolein. Each one of these compounds is itself a secondary product of hydrocarbon oxidation. Of these, acetaldehyde is likely to be the most important PA precursor. The photochemical sink of PAN is through thermal dissociation followed by reaction of PA with NO (ref. 5). During the summer day these reactions are rapid and the PAN mixing ratio tends to follow the mixing ratio of NO₂. At night, without the presence of appreciable OH and NO, the photochemical sources and sinks of PAN are interrupted, and the PAN concentration is controlled by surface deposition and atmospheric mixing. During the autumn day, the OH levels are sufficiently reduced that PAN production is effectively suppressed while the lower temperature reduces the thermal decomposition of PAN. Consequently, the PAN levels during the autumn day are relatively independent of NO_x.

It is possible to draw more quantitative conclusions from the summer data. The abundance of acetaldehyde and PA radicals that are needed to explain the observed [PAN] values can be estimated by assuming photochemical equilibrium for PA and PAN. For the summer afternoon with relatively clean air an 'apparent' acetaldehyde mixing ratio of 150–300 p.p.t.v. and PA radical concentrations of $2\text{--}3 \times 10^7$ molecules cm⁻³ are consistent with our observations. Note that the 'apparent' acetaldehyde mixing ratio deduced is an upper limit for the actual acetaldehyde mixing ratio because PA radicals can also be formed by other paths. No measurements of acetaldehyde from remote sites are available. However, in urban/suburban environments, acetaldehyde mixing ratios in the range of 0.2–20 p.p.b.v. (refs 17–19) have been reported.

From these concurrent measurements of PAN, PPN, NO_x, O₃ and meteorological parameters at a remote site in the

Colorado mountains we conclude that nitrogen oxides present in the form of PAN can be as much or more abundant than the inorganic form. The importance of PAN relative to NO_x increases with decreasing values of NO_x as well as with decreasing temperatures. It is suggested that only a modest amount of acetaldehyde and other precursors is needed to account for this PAN synthesis. The photo-oxidation of non-methane hydrocarbons probably provides these precursors in the background air. Extrapolation of these results to the upper troposphere, for which no simultaneous PAN-NO_x data are currently available, implies a highly significant role for PAN and other organic nitrates in the transport of combined nitrogen through the atmosphere. In addition, PAN can be an important source of peroxyacetyl radicals which may be important to oxidation processes in the gas as well as liquid phases.

The SRI International study is funded by NSF under grant ATM-8308816. The National Center for Atmospheric Research is sponsored by the NSF. The Aeronomy Laboratory research is supported primarily by the National Oceanic and Atmospheric Administration as part of the National Acid Precipitation Assessment Program.

Received 17 June; accepted 4 September 1985.

1. Haagen-Smit, A. J. *Indust. Engng Chem.* **44**, 1342–1346 (1952).
2. Chameides, W. L. *Geophys. Res. Lett.* **5**, 17–20 (1978).
3. Crutzen, P. J. *A. Rev. Earth planet. Sci.* **7**, 443–472 (1979).
4. Logan, J. A., Prather, M. J., Wofsy, S. C. & McElroy, M. B. *J. geophys. Res.* **86**, 7210–7254 (1981).
5. Singh, H. B. & Hanst, P. L. *Geophys. Res. Lett.* **8**, 941–944 (1981).
6. Aikin, A. C., Herman, J. R., Maier, E. J. & McQuillan, C. J. *J. geophys. Res.* **87**, 3105–3118 (1982).
7. Singh, H. B. & Salas, L. J. *Nature* **302**, 326–328 (1983).
8. Spicer, C. W., Holdren, M. W. & Keigley, G. W. *Atmos. Envir.* **17**, 1055–1057 (1983).
9. Bollinger, M. J. thesis, Univ. Colorado (1982).
10. Bollinger, M. J. *et al. J. geophys. Res.* **89**, 9623–9631 (1984).
11. Singh, H. B. & Salas, L. J. *Atmos. Envir.* **17**, 1507–1516 (1983).
12. Lewis, T. E., Brennan, E. & Lonneman, W. A. *JAPCA* **33**, 885–887 (1983).
13. Lonneman, W. A., Bufalini, J. J. & Seila, R. L. *Envir. Sci. Technol.* **10**, 374–380 (1976).
14. Brice, K. A. *et al. Atmos. Envir.* **18**, 2691–2702 (1984).
15. Fehsenfeld, F. C. *et al. J. atmos. Chem.* **1**, 87–105 (1983).
16. Parrish, D. D. *et al. J. atmos. Chem.* (submitted).
17. Tanner, R. L. & Meng, Z. *Envir. Sci. Technol.* **18**, 723–726 (1984).
18. Singh, H. B., Salas, L. J., Stiles, R. & Shigeishi, H. *Measurements of Hazardous Organic Chemicals in the Ambient Air* (NTIS Rep. PB83-156935, 1983).
19. Grosjean, D. *Envir. Sci. Technol.* **16**, 254–262 (1982).

Isotopic composition of atmospheric O₂ in ice linked with deglaciation and global primary productivity

M. Bender*†, L. D. Labeyrie*, D. Raynaud† & C. Lorius†

* Centre des Faibles Radioactivites, CNRS-CEA, BP no. 1—91190, Gif/Yvette, France

† Laboratoire de Glaciologie et Geophysique de l'Environnement, BP 96-38402 St Martin d'Heres Cedex, France

In photosynthesis, O₂ is continuously formed from H₂O and released to the atmosphere. Coupled with respiration, photosynthesis forms a loop in which oxygen isotopes are exchanged between O₂ and H₂O. During the ice ages, sea water was enriched in $\delta^{18}\text{O}$ by $\sim 1.3\%$ relative to the present value¹. Continental waters in the areas of high primary productivity exchange rapidly with the oceans. They probably presented a similar isotopic enrichment. Since the $\delta^{18}\text{O}$ of glacial water was greater than at present, we would expect that the $\delta^{18}\text{O}$ of atmospheric O₂ was also greater than at present. Fireman and Norris² and Horibe *et al.*³ have measured the $\delta^{18}\text{O}$ of O₂ from the glacial atmosphere by analysing trapped gases in ice cores. However, their data are either too few or too imprecise to demonstrate whether $\delta^{18}\text{O}$ of atmospheric O₂

† Present address: Graduate School of Oceanography, University of Rhode Island, Kingston, Rhode Island 02881, USA.

has, in fact, varied. Here we present data on the changes, during the past 22 kyr approximately, in the $\delta^{18}\text{O}$ of atmospheric O_2 trapped in the ice core Dome C (East Antarctica, 74°S , 124°E). The results show that the isotopic composition of atmospheric O_2 has indeed varied along with that of sea water, and that the $\delta^{18}\text{O}$ (O_2) record offers a tool for studying several important aspects of the global cycles of O_2 and H_2O in relation to the climate.

When the snow deposited at the surface of ice sheets transforms into ice, atmospheric air is trapped in the form of bubbles. The analysis of this fossil air extracted from ice cores has already provided spectacular information, especially on the CO_2 composition of the atmosphere during the past 40 kyr (refs 4, 5). The reliability of the measurements of the composition of trapped air depends on the absence of melting during the formation and diagenesis of the solid H_2O particles, and the lack of alteration of air bubble composition during and after sampling⁶. Based on the reliable record of atmospheric CO_2 measured on the Dome C ice core⁴, we expect that it is one of the best existing cores for determining with confidence the $\delta^{18}\text{O}$ of atmospheric oxygen during the past 25 kyr. Furthermore, the special interest of this core is due to the fact that it contains a reliable high-latitude climate record, which is reflected in the $\delta^{18}\text{O}$ of the ice⁷.

To determine the $\delta^{18}\text{O}$ of O_2 , ice samples weighing $\sim 60\text{ g}$ were placed in glass flasks and evacuated to $1\text{ }\mu\text{atm}$ pressure. Trapped air was extracted from ice by melting and slowly refreezing the water, using an experimental procedure similar to the melting method described in ref. 8. The O_2 in the extracted air was converted to CO_2 by exposure for 40 min to a hot graphite rod⁹. At the end of the combustion, water was separated from CO_2 using a frozen isopropyl alcohol trap, and the water vapour pressure was measured. The $\delta^{18}\text{O}$ of the CO_2 was determined using a Micromass 602D mass spectrometer. We found that ^{18}O fractionation during combustion is geometry-dependent and not, in general, as large as values measured by Kroopnick¹⁰. The efficiencies of air extraction and O_2 conversion to CO_2 (checked by measuring the O_2/N_2 ratio in the residual gas after combustion) were estimated to be >99 and $>99.85\%$, respectively. We did not measure residual levels of CO , which may have been formed during O_2 conversion and caused our poor reproducibility for some samples. Excluded from the compilation of our data here and in Table 1 are results from 14 samples and air standards for which the water vapour pressure after combustion was $>10\text{ }\mu\text{atm}$, the carbon rod was heated to an

anomalously high temperature, or the combustion was terminated before the standard 40-min interval. Results from nine of the excluded samples are in good agreement with data presented here. The $\delta^{18}\text{O}$ of modern atmospheric O_2 , sampled in Gif and determined using the same procedure as for ice samples, is $+23.81 \pm 0.31\%$ ($n=9$ air samples) where $\delta^{18}\text{O} = 10^3 \times ([^{18}\text{O}/^{16}\text{O}]_{\text{sample}} - [^{18}\text{O}/^{16}\text{O}]_{\text{standard}}) / ([^{18}\text{O}/^{16}\text{O}]_{\text{standard}})$, the standard being standard mean ocean water (SMOW)¹¹. This value of $23.81 \pm 0.31\%$ is in satisfactory agreement with the present atmospheric value of $+23.5\%$ determined by Kroopnick and Craig¹². Our measurements on three postglacial ice samples from Antarctic ice core D10 have a mean $\delta^{18}\text{O}(\text{O}_2) = 23.68 \pm 0.26\%$, and similar values were measured for postglacial Dome C samples (Table 1). These results suggest that our extraction procedure does not fractionate isotopically the oxygen initially trapped in the air bubbles of the ice and, furthermore, that this O_2 is an isotopically unfractionated sample of the contemporaneous atmosphere. Measured $\delta^{18}\text{O}$ values of past atmospheric O_2 are given in Table 1 and Fig. 1c.

Data on the $\delta^{18}\text{O}$ of the ice in ice cores, along with that of benthic foraminifera from deep-sea sediments, provide a context for interpreting the $\delta^{18}\text{O}$ of palaeoatmospheric O_2 . Normalized $\delta^{18}\text{O}$ values of benthic foraminifera in the Indian Ocean core MD73025 (ref. 13) are shown as a function of age in Fig. 1b. The 1.6% decrease in $\delta^{18}\text{O}$ between 16 and 6 kyr BP mainly reflects the addition to the ocean of glacial meltwater depleted in ^{18}O (ref. 1), although a fraction of the signal, $\sim 0.3\%$, is probably due to a warming of the deep Antarctic circumpolar waters¹⁴. The $\delta^{18}\text{O}$ decrease associated solely with continental ice melting is believed to be $1.3 \pm 0.1\%$ (ref. 14). The $\delta^{18}\text{O}$ of ice in the Dome C ice core is shown as a function of age in Fig. 1a. The sharp increase, going towards the present, reflects the marked warming at high latitudes associated with the last glacial termination. The timescale for the ice has been obtained using a simple ice flow model⁷, assuming a variable accumulation rate related to temperature changes, as suggested by comparison with well-dated events inferred from the foraminiferal $\delta^{18}\text{O}$ profile of core MD73025 (ref. 13).

$\delta^{18}\text{O}$ of O_2 gas in the Dome C core is plotted in Fig. 1c. At a given depth gases in ice are younger than ice itself because of the time required for snow to be transformed into ice and seal the bubbles⁶. According to Schwander and Stauffer¹⁵, 80% of the bubbles close off between the snow densities 0.795 and 0.830. The mean ages of the gas have been calculated assuming

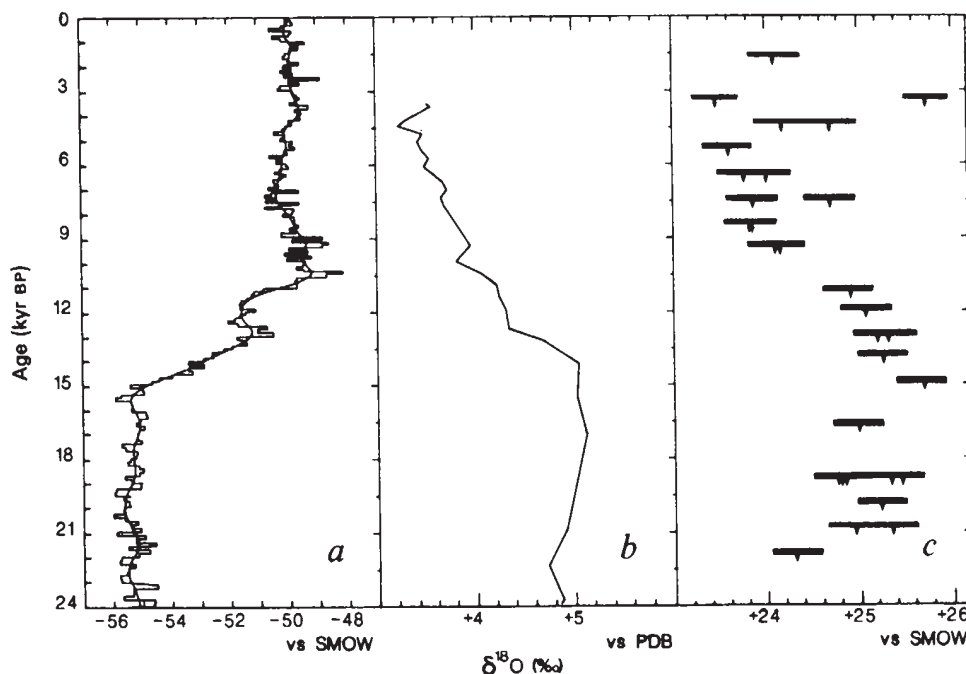


Fig. 1 a, $\delta^{18}\text{O}$ of ice versus age of the ice. This plot is identical to that in ref. 7 except that the ordinate here is linear with age. b, $\delta^{18}\text{O}$ of benthic foraminifera in core MD73025 versus age. c, $\delta^{18}\text{O}$ of O_2 gas in air trapped in Dome C versus mean age of the gas. The horizontal bars correspond to the analytical uncertainty of $\pm 0.25\%$, and the arrows indicate that the mean age of the gas may be older by a few hundred years.

Table 1 $\delta^{18}\text{O}$ of O_2 in air trapped at different depths in Dome C, along with ice equivalent depth, age of the ice and estimated mean age of the gas at each level

Depth below the surface (m)	Depth in ice equivalent (m)	Age of the ice (kyr)	Mean age of the gas (kyr)	$\delta^{18}\text{O}$ of the O_2 (% vs SMOW)
150.9	118.6	3.26	1.56	24.04
210.8	179.1	4.97	3.27	25.61
210.8	179.1	4.97	3.27	23.45
246.2	215.1	6.00	4.30	24.10
246.2	215.1	6.00	4.30	24.62
279.2	248.1	6.96	5.26	23.57
315.6	283.5	8.03	6.33	23.70
315.6	283.5	8.03	6.33	23.97
350.2	319.1	9.06	7.36	24.62
350.2	319.1	9.06	7.36	23.82
384.3	353.2	10.08	8.38	23.78
384.3	353.2	10.08	8.38	23.80
414.4	383.3	10.98	9.27	24.05
414.4	383.3	10.98	9.27	24.09
476.9	445.8	13.03	11.09	24.81
500.8	469.7	13.91	11.86	24.97
530.4	499.3	15.07	12.87	25.21
530.4	499.3	15.07	12.87	25.12
552.3	521.2	15.99	13.73	25.14
578.3	547.2	17.10	14.84	25.54
578.3	547.2	17.10	14.84	25.02
618.1	587.1	18.82	16.56	24.89
667.5	636.5	20.99	18.73	24.68
667.5	636.5	20.99	18.73	25.17
667.5	636.5	20.99	18.73	25.31
667.5	636.5	20.99	18.73	24.74
667.5	636.5	20.99	18.73	24.77
690.7	659.7	22.02	19.76	25.13
712.7	681.6	23.01	20.75	24.83
712.7	681.6	23.01	20.75	25.23
736.4	705.3	24.08	21.81	24.24

that the bubbles close off at a constant rate between the densities 0.8 and 0.83. The corresponding mean age values are given in Table 1. However, different factors discussed in ref. 16 may accelerate the close off, and the mean ages of the gas may, therefore, be older by a few hundred years in the case of Dome C. This effect is suggested by the arrows in Fig. 1. Two important results emerge from the $\delta^{18}\text{O}$ (O_2)/age curve (Fig. 1c). First, during the last glacial period, the mean $\delta^{18}\text{O}$ of atmospheric O_2 was higher than the present value by 1.3%. Second, the change in the $\delta^{18}\text{O}$ of atmospheric O_2 and that of sea water (inferred from foraminiferal $\delta^{18}\text{O}$) occurred approximately simultaneously, given the uncertainties in dating procedures and scatter in the data.

Concerning the first point, the average $\delta^{18}\text{O}$ of samples younger than 7 kyr is $+23.76 \pm 0.25\%$ ($n = 5$) whereas the average $\delta^{18}\text{O}$ for samples older than 15 kyr is $+25.06 \pm 0.24\%$ ($n = 14$). Anomalous results on samples from depths of 211 and 736 m have been excluded in this averaging. Their inclusion increases the standard deviation but does not significantly change the mean values (giving $+23.90\%$ for samples younger than 7 kyr, and $+25.00\%$ for samples older than 15 kyr). Our results thus confirm the expectation discussed earlier, that the $\delta^{18}\text{O}$ of atmospheric O_2 was heavier during the last glacial than it is today. We believe that we can dismiss the possibility that this glacial/postglacial $\delta^{18}\text{O}(\text{O}_2)$ difference is due to oxygen isotope exchange in the ice. First, no inorganic reactions between H_2O and O_2 are known which occur in the absence of light at low temperatures. Second, with an ice $\delta^{18}\text{O}$ around -50% and a fractionation factor at equilibrium of $\sim 20\%$ (ref. 17), the $\delta^{18}\text{O}$ of trapped O_2 approaching isotopic equilibrium with ice would decrease progressively during ageing (to about -30%). The opposite effect is observed.

The change in the $\delta^{18}\text{O}$ of atmospheric O_2 between the last glacial and the present (1.3%) is identical to the corresponding

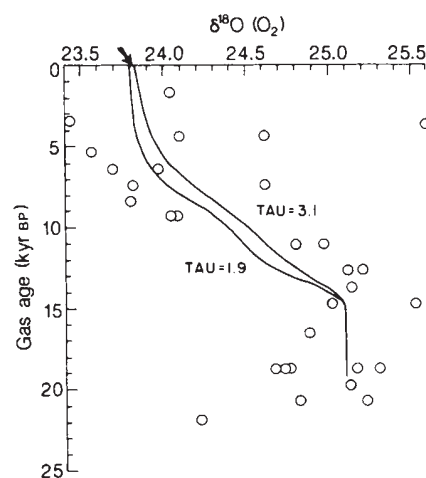


Fig. 2 Model values of atmospheric $\delta^{18}\text{O}$ (O_2) versus age BP, calculated for atmospheric O_2 turnover times of 1.9 and 3.1 kyr, along with $\delta^{18}\text{O}$ (O_2) values measured in the Dome C ice core. The modern atmospheric value is shown by the arrow. The calculations assume melting, at a constant rate, of 60% of all ice which melted during the last glacial termination, between 15.0 and 13.0 kyr BP, the remainder melting at a constant rate between 10.5 and 8 kyr BP. Surface sea water $\delta^{18}\text{O}$ is taken as 0‰ at present, and $+1.3\%$ before 15 kyr BP; it is assumed to vary proportionately to ice volume during intermediate times. Model calculations are done using the equation: $d\delta^{18}\text{O}(t)/dt = (\delta^{18}\text{O}(\text{ss}) - \delta^{18}\text{O}(t))/\text{TAU}$, where t is time, TAU is atmospheric O_2 turnover time, and $\delta^{18}\text{O}(\text{ss})$ and $\delta^{18}\text{O}(t)$ are the $\delta^{18}\text{O}$ values of atmospheric O_2 at steady state and at time t , respectively. The steady-state value is taken as $+23.5\%$ with respect to surface sea water.

change in sea water $\delta^{18}\text{O}$. This similarity may be surprising, as can be seen from considering the controls on the $\delta^{18}\text{O}$ of atmospheric O_2 . Isotopic exchange between water and atmospheric oxygen occurs through transfer of oxygen atoms within the biological pathways, by photosynthesis and respiration. At steady state, the $\delta^{18}\text{O}$ of O_2 consumed by respiration must be equal to the $\delta^{18}\text{O}$ of O_2 produced by photosynthesis. The $\delta^{18}\text{O}$ of photosynthetic O_2 is: $\delta^{18}\text{O}(\text{photosynthetic } \text{O}_2) = \delta^{18}\text{O}(\text{water}) + A$, where $\delta^{18}\text{O}(\text{water}) = \delta^{18}\text{O}(\text{ocean}) + W$, A equals the kinetic isotope effect during photosynthesis¹⁸ and W is the weighted mean $\delta^{18}\text{O}$ difference between the global sea water $\delta^{18}\text{O}$ and the $\delta^{18}\text{O}$ of the water which is the immediate source for photosynthetic oxygen. The $\delta^{18}\text{O}$ of respiratory O_2 is: $\delta^{18}\text{O}(\text{respiratory } \text{O}_2) = \delta^{18}\text{O}(\text{O}_2) + B$, where $\delta^{18}\text{O}(\text{O}_2)$ is the atmospheric O_2 $\delta^{18}\text{O}$ and B is the respiratory kinetic isotope effect^{19,20}. At steady state, $\delta^{18}\text{O}(\text{O}_2) - \delta^{18}\text{O}(\text{ocean}) = W + A - B$. All three terms on the right-hand side probably changed between glacial and interglacial times. W varied for terrestrial photosynthesis because of changing isotope effects in the hydrological cycle and in evapotranspiration. A and B (the kinetic photosynthetic and respiratory isotope effects) varied because different plants, animals and bacteria were responsible for O_2 production and consumption during glacial times. The similarity of the $(W + A - B)$ term between the last ice age and today, as inferred from our data, may reflect a very strong stability in oxygen isotope fractionation by the most productive marine and terrestrial ecosystems, and the hydrological regimes in which they lie.

The second interesting feature of the $\delta^{18}\text{O}(\text{O}_2)$ /age record concerns the shape of the curve during the last glacial termination. The rate at which the glacial steady-state O_2 $\delta^{18}\text{O}$ value ($+25.1\%$) approaches the postglacial value ($+23.8\%$) depends on the time course of change in the $\delta^{18}\text{O}$ of sea water due to deglaciation, and the lag in the isotopic equilibration of atmospheric O_2 with respect to photosynthesis. The latter term is, in turn, equal to the magnitude of the global O_2 reservoir divided by the gross rate of photosynthesis on the planet. The atmospheric O_2 turnover time is estimated to be in the range of 3.1–1.9 kyr (Table 2).

Table 2 Parameters for estimating the turnover time of atmospheric O₂

Global atmospheric O ₂ reservoir (GO ₂ R) = 3.7×10^{19} mol
Terrestrial net primary productivity (TNPP) = 5×10^{15} mol yr ⁻¹ (refs 21, 22)
Terrestrial gross primary productivity (TGPP) = $2 \times$ TNPP (ref. 23) = 10×10^{15} mol yr ⁻¹
Marine primary productivity (MPP) = 2×10^{15} (ref. 24)– 10×10^{15} mol yr ⁻¹ (refs 24, 25)
Global primary productivity (GPP) = (TGPP + MPP) = 12×10^{15} – 20×10^{15} mol yr ⁻¹
Atmospheric O ₂ turnover time = (GO ₂ R/GPP) = 3.1–1.9 kyr

In Fig. 2, results are given for simple model calculations of $\delta^{18}\text{O}$ (O₂) versus time (described in Fig. 2 legend), for the atmospheric O₂ turnover times estimated above. The comparison of the model curves with the observed variation in $\delta^{18}\text{O}$ of atmospheric O₂ suggests a lower limit on global productivity around the time of the last glacial termination: our data, although scattered, show that the decrease in atmospheric $\delta^{18}\text{O}$ to the postglacial values, initiated at the earliest around 13 kyr BP, was nearly complete by 9 kyr BP. The oxygen isotopic decrease was, therefore, at least as fast as estimated from the model, assuming a turnover time of 3.1 kyr (Fig. 2). This result suggests that productivity values during deglaciation were $\geq 12 \times 10^{15}$ mol yr⁻¹ (the low estimate for present productivity). This does not exclude the possibility that global primary productivity during the last glacial termination was less than at present, if current productivity is on the high side of the different estimates.

We conclude that, during the past 22 kyr, the $\delta^{18}\text{O}$ of atmospheric O₂ has changed along with that of sea water. The change reflects, in a complex way, the controls on the oxygen isotope fractionation between sea water and atmospheric O₂ and the rate of global primary productivity. In addition, since atmospheric O₂ is well mixed isotopically, $\delta^{18}\text{O}$ (O₂) is a time-stratigraphical marker. It can be used for the correlation between climatic records obtained from marine and ice cores. This is specially important for interpreting palaeo-CO₂ records from ice.

We acknowledge discussions with J. Labeyrie, C. Lalou and J. C. Duplessy. This work was done with support from Centre National de la Recherche Scientifique, Commissariat à l'Energie Atomique, and Programme National d'Etude de la Dynamique du Climat. M.B. acknowledges hospitality and support from CFR and CNRS as a visiting scientist. The Dome C ice core collection was funded by Expeditions Polaires Françaises, Terres Australes et Antarctiques Françaises, and US-NSF (Division of Polar Program).

Received 27 June; accepted 19 September 1985.

- Shackleton, N. J. *Nature* **215**, 15–17 (1967).
- Fireman, E. L. & Norris, T. L. *Earth planet. Sci. Lett.* **60**, 339–350 (1982).
- Horibe, Y., Shigehara, K. & Langway, C. Jr *Earth planet. Sci. Lett.* **73**, 207–210 (1985).
- Delmas, R. J., Ascencio, J. M. & Legrand, M. *Nature* **284**, 155–157 (1980).
- Neffel, A., Oeschger, H., Schwander, J., Stauffer, R. & Zumbun, R. *Nature* **295**, 220–223 (1982).
- Lorius, C. & Raynaud, D. in *Carbon Dioxide: Current Views and Developments in Energy/Climate Research* (eds. Bach, W. et al.) 145–176 (Riedel, Dordrecht, 1983).
- Lorius, C., Merlivat, L., Jouzel, J. & Pourchet, M. *Nature* **280**, 644–648 (1979).
- Raynaud, D., Delmas, R., Ascencio, J. M. & Legrand, M. *Ann. Glaciol.* **3**, 265–268 (1982).
- Labeyrie, L. D. *Note CEA-N-2086*, CEN Saclay, France (1979).
- Kroopnick, P. thesis, Univ. California, San Diego (1970).
- Craig, H. *Geochim. cosmochim. Acta* **12**, 133–149 (1957).
- Kroopnick, P. & Craig, H. *Science* **175**, 54–55 (1972).
- Duplessy, J. C., Moyes, J. & Pujol, C. *Nature* **286**, 479–482 (1980).
- Duplessy, J. C., Labeyrie, L. & Shackleton, N. J. *Eos* **66**, 292 (1985).
- Schwander, J. & Stauffer, B. *Nature* **311**, 45–47 (1984).
- Raynaud, D. & Barnola, J. M. *Nature* **315**, 309–311 (1985).
- Urey, H. C. *J. chem. Soc.* 562–581 (1947).
- Stevens, C. L. R., Schultz, D., Van Baalen, C. & Parker, P. L. *Pl. Physiol.* **56**, 126–129 (1975).
- Lane, G. A. & Dole, M. *Science* **123**, 574–576 (1966).
- Kroopnick, P. & Craig, H. *Earth planet. Sci. Lett.* **32**, 375–388 (1976).
- Ajtay, G. L., Ketner, P. & Duvigneaud, P. in *The Global Carbon Cycle (SCOPE 13)* (eds Bolin, B., Degens, E. T., Kempe, S. & Ketner, P.) 129–182 (Wiley, New York, 1979).
- Woodwell, G. M. et al. *Science* **199**, 141–146 (1978).
- Bramryd, T. in *The Global Carbon Cycle (SCOPE 13)* (eds Bolin, B., Degens, E. T., Kempe, S. & Ketner, P.) 183–218 (Wiley, New York, 1979).
- Eppley, R. W. in *Primary Productivity in the Sea* (ed. Falkowski, P.) 230–242 (Plenum, New York, 1980).
- Jenkins, W. J. *Nature* **300**, 246–248 (1982).

Silicon coordination changes from 4-fold to 6-fold on devitrification of silicon phosphate glass

T. L. Weeding*, B. H. W. S. de Jong†, W. S. Veeman* & B. G. Aitken‡

* Department of Physical Chemistry, Faculty of Science, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands
† Smitweld, B.V., PO Box 253, 6500 AG Nijmegen, The Netherlands
‡ Corning Glass Works, Corning, New York 14831, USA

4-Fold oxygen coordinated tetrahedral silicon (^{IV}Si) is the common building block in silicas and silicates. However, 17 crystallographically well-characterized inorganic silicates are known in which silicon exists in 6-fold oxygen octahedral coordination¹. The most notable is stishovite, a high-pressure SiO₂ polymorph^{2,3}. An interesting question is whether differences in silicon oxygen coordination between liquid and solid silicates impose a barrier to nucleation. One such case is stishovite (^{VI}Si) which cannot be synthesized from SiO₂ with ^{IV}Si at ambient pressures^{4,5}. That such an apparent nucleation barrier is not limited to silicon oxygen coordination is shown by aluminium. Thus, jadeite (NaAlSi₃O₆) with ^{VI}Al will not precipitate from its isochemical glass, in which only ^{IV}Al has been proven to exist, except when subjected to pressures >60 kbar^{6,7}. To test if transitions of silicon coordination from 4- to 6-fold impose, in general, a barrier to nucleation, we studied the SiO₂–P₂O₅ system. X-ray diffraction studies on one of the phases in this system, SiO₂·P₂O₅, show the presence of ^{VI}Si only⁸. The aforementioned examples indicate that the silicon-phosphate glass from which this phase precipitates will also contain ^{VI}Si, as no external pressure is used. Zachariassen⁹, on the other hand, has argued against 6-fold coordination in glasses, predicting that octahedral coordination would force periodicity on the lattice, thereby disrupting the vitreous state. Our ²⁹Si MAS (magic angle spinning) NMR results show that the SiO₂–P₂O₅ glass contains only ^{IV}Si which on devitrification at ambient pressures transforms to crystalline silicon-phosphate with ^{VI}Si.

The SiO₂–P₂O₅ glasses were prepared at Corning Glass Works (Corning, New York). The glasses, which were shown to be X-ray amorphous, were devitrified at 850, 1,000 and 1,300 °C. The nature of the crystalline constituents were determined by X-ray diffraction. The X-ray pattern of the sample devitrified at 1,300 °C showed only crystals of the cubic phase of SiP₂O₇ on top of an amorphous halo. The samples devitrified at 1,000 and 850 °C contained crystals of Si₅O(PO₄)₆, SiP₂O₇ and an amorphous component¹⁰.

A single-crystal X-ray study by Tillmans *et al.*⁸ has confirmed the 6-fold coordination about Si in the cubic polymorph of SiP₂O₇ determined by earlier workers^{11,12}. They found six crystallographically distinguishable silicon sites with Si–O bonds of different lengths and a range of Si–O–P angles. All of the phosphorus exists as pyrophosphate. Si₅O(PO₄)₆ consists of isolated SiO₆ octahedra and Si₂O₇ groups with ^{IV}Si which are linked by PO₄ tetrahedra¹³. The phase diagram for this SiO₂–P₂O₅ system has been reported by Tien and Hummel¹⁴. (However, the 5SiO₂·3P₂O₅ component was incorrectly assigned^{13,15} as 2SiO₂·P₂O₅.)

²⁹Si-NMR data were collected on a Bruker CXP-300 spectrometer operating at 59.6 MHz with a MAS accessory. The magic angle was adjusted with benitoite¹⁶. The spinning rate was 2.5 kHz using an Andrew-type spinner. The chemical environments of the phosphorus atoms were also examined using ³¹P MAS NMR on a home-built 180-MHz spectrometer¹⁷ operating at 72.854 MHz.

Figure 1 shows ²⁹Si MAS NMR spectra obtained from a glass and from the same glass devitrified at the three different temperatures. The lines near –110 p.p.m. (parts per 10⁶) are characteristic of tetrahedrally coordinated Si and those around

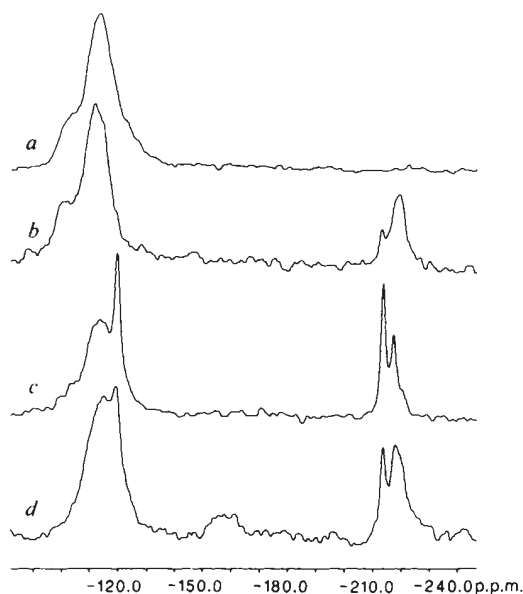


Fig. 1 ^{29}Si MAS NMR spectra from the $\text{SiO}_2\text{-P}_2\text{O}_5$ system. The chemical shift is referenced to a quartz external standard (-107.5 p.p.m.). Excitation pulses were $4\text{ }\mu\text{s}$ long, corresponding to $\sim 60^\circ$ pulses. Delay between pulses was 20 s for the glass and 1 s for the devitrified samples. Increasing the delay up to 10 min for the glass and to 5 min for the devitrified samples did increase the intensity of the octahedral lines in the devitrified samples by $\sim 25\%$ but no octahedral signal was observed in the glass. *a*, Spectrum of the glass resulting from $12,000$ FID (free induction decays). *b*, Spectrum of glass devitrified at $1,300^\circ\text{C}$. The X-ray powder pattern showed the presence of SiP_2O_7 only. The spectrum is from $72,250$ FID. *c*, Glass devitrified at $1,000^\circ\text{C}$ for 6 h . The X-ray powder pattern showed the presence of both SiP_2O_7 and $\text{Si}_5\text{O}(\text{PO}_4)_6$, the latter in the greater amount. $69,700$ FID. *d*, Spectrum of glass devitrified at 850°C for 16 h . The X-ray powder pattern showed that crystals of SiP_2O_7 and $\text{Si}_5\text{O}(\text{PO}_4)_6$ were present in similar amounts. $12,750$ FID were co-added and transformed to give the spectrum.

-215 p.p.m. indicate the presence of the octahedral coordination¹⁸. The spectrum of the glass, shown in Fig. 1*a*, and the spectra of the three different partially crystalline samples, shown in Fig. 1*b-d*, demonstrate that silicon exhibits only 4-fold coordination in the glass but, on devitrification, 6-fold coordinated silicon appears.

The X-ray powder pattern of the sample devitrified at $1,000^\circ\text{C}$ showed twice the ratio of $\text{Si}_5\text{O}(\text{PO}_4)_6\text{:SiP}_2\text{O}_7$ that was seen in the sample devitrified at 850°C . Hence, a comparison between the spectrum in Fig. 1*b* containing only SiP_2O_7 and those in Fig. 1*c* and *d*, which contain different relative amounts of both crystalline components, allows assignment of the peaks. The SiP_2O_7 (Fig. 1*b*) shows a narrow octahedral line at -214 ± 1 p.p.m. and a broad line centred at -220 p.p.m. In Fig. 1*c* and *d*, the transition at -214 ± 1 p.p.m. changes with composition, due to $\text{Si}_5\text{O}(\text{PO}_4)_6$ having ^{31}Si . The $\text{Si}_5\text{O}(\text{PO}_4)_6$ probably also contributes to the intensity at -217 ± 1 p.p.m. These chemical shift values may be contrasted with the value of -191 p.p.m. for stishovite¹⁸ and -179.5 p.p.m. for thaumasite¹⁹, the only solids containing ^{31}Si for which ^{29}Si NMR data have been collected to date. The narrow peak in Fig. 1*c* and *d* at -120 ± 1 p.p.m. is assigned to ^{31}Si in the crystalline $\text{Si}_5\text{O}(\text{PO}_4)_6$ phase.

A substantial chemical shift of the ^{31}Si occurs relative to usual Q_1 shifts: compare ^{31}Si in $\text{Si}_5\text{O}(\text{PO}_4)_6$ with a O-Si-O angle of 180° and chemical shift of ~ -120 p.p.m. with ^{31}Si in $\text{Sc}_2\text{Si}_2\text{O}_7$ also with a O-Si-O angle of 180° and chemical shift of -93.5 p.p.m. (ref. 1). This upfield shift is attributed to phosphorus. The relative intensity of the amorphous peak that is the most downfield (-103 p.p.m.) decreases more rapidly with the progress of devitrification than does the upfield peak.

The ^{31}P -NMR spectra reflect several different environments for the phosphorus nuclei. For the glass, we see five lines with chemical shifts of 15 , -0.4 , -12 , -54 and -72 p.p.m. relative to an external standard of 85% H_3PO_4 . The partially devitrified samples showed changes in the intensity of these lines and, in the multi-component cases, a new line at -44 p.p.m. which must be due to the crystalline $\text{Si}_5\text{O}(\text{PO}_4)_6$ phase. The spectra are complicated by overlapping spinning side bands due to the large chemical shift anisotropy of ^{31}P . Assignment of the peaks is also difficult due to the range of chemical shifts which have been reported for phosphates^{20,21}.

The reconstructive transitions, requiring the making and breaking of strong Si-O bonds, which change the coordination of the Si from 4-fold in the glass to 6-fold in the crystal occurs in this system at ambient pressures. The 4 to 6-fold transition does not prohibit nucleation of the crystalline phase, indeed, the phase containing only ^{31}Si appears to nucleate first. The presence of only tetrahedral coordination in the glass is consistent with Zachariasen's prediction.

This work was supported by the Dutch Foundation for Pure Research (ZWO/SON).

Received 24 June; accepted 22 August 1985.

1. Liebau, F. *Structural Chemistry of Silicates* 347 (Springer, Berlin, 1985).
2. Baur, W. H. & Khan, A. A. *Acta crystallogr.* **B27**, 2133-2139 (1977).
3. Kirkpatrick, R. J., Smith, K. A., Schramm, S., Turner, G. & Yang, W.-H. A. *Rev. Earth planet. Sci.* **13**, 29-47 (1985).
4. Stishov, S. M. & Popova, S. V. *Geokhimiya* **10**, 837-839 (1961).
5. Frondel, C. *The System of Mineralogy* Vol. 3 (Wiley, New York, 1962).
6. De Jong, B. H. W. S., Schramm, C. M. & Parziale, V. E. *Geochim. cosmochim. Acta* **47**, 1223-1236 (1983).
7. Ohtani, E., Taulelle, F. & Angell, C. A. *Nature* **314**, 78-81 (1985).
8. Tillmans, E., Gebert, W. & Baur, W. H. *J. Solid State Chem.* **7**, 69-84 (1973).
9. Zachariasen, W. H. *J. Am. chem. Soc.* **54**, 3841-3851 (1932).
10. Makart, H. *Helv. chim. Acta* **50**, 399-405 (1967).
11. Levi, G. R. & Peyronel, G. Z. *Kristallogr.* **92**, 190-209 (1935).
12. Vollenkle, H., Wittman, A. & Nowotny, H. *Mh. Chem.* **94**, 956-963 (1963).
13. Mayer, H. *Mh. Chem.* **105**, 46-54 (1974).
14. Tien, T. Y. & Hummel, F. A. *J. Am. Ceram. Soc.* **45**, 422-424 (1962).
15. Liebau, F., Bissert, G. & Koppen, N. *Z. anorg. allg. Chem.* **359**, 131-134 (1968).
16. De Jong, B. H. W. S., Schramm, C. M. & Parziale, V. E. *J. Am. chem. Soc.* **106**, 4396-4402 (1984).
17. Veeman, W. S., Menger, E. M., Ritchey, W. & de Boer, E. *Macromolecules* **12**, 924-927 (1979).
18. Smith, J. W. & Blackwell, C. S. *Nature* **303**, 223-225 (1983).
19. Grimmer, A. E. *et al.* *J. Chem.* **20**, 453 (1980).
20. Tropp, J., Blumenthal, N. C. & Waugh, J. S. *J. Am. chem. Soc.* **105**, 22-26 (1983).
21. Van Wazer, J. R., Callis, C. F., Shoolery, J. N. & Jones, R. C. *J. Am. chem. Soc.* **78**, 5715-5726 (1956).

Transfer of structural information from Langmuir monolayers to three-dimensional growing crystals

Ehud M. Landau*, Moshe Levanon*, Leslie Leiserowitz*, Meir Lahav* & Jacob Sagiv†

Departments of Structural Chemistry* and Isotope Research†, Weizmann Institute of Science, Rehovot 76100, Israel

One enantiomer of an initially racemic mixture of α -amino acids can be preferentially removed from solution by selective incorporation into the opposite enantiotopic faces of growing centrosymmetric crystals of the α -form of glycine^{1,16}. We report here that oriented growth of crystals of the α form of glycine has been achieved under chiral Langmuir monolayers comprising amphiphilic α -amino acids, by virtue of a structural match between the monolayer and the *ac* surface layer of the attached growing glycine crystals. Such monolayers of α -amino acid of *R* configuration, containing long hydrocarbon chains, induce glycine to crystallize with its (010) face attached to the monolayer, and by symmetry the corresponding *S* amino-acid monolayers induce attachment of the (010) face of glycine. Replacement of the hydrocarbon by a fluorocarbon chain induces analogous crystallizations, albeit, with only a partial degree of orientation, whereas monolayers of a resolved amino acid bearing a cholestanoyl moiety do not promote crystallization of glycine. Monolayers of the *R*, *S*-amino acids

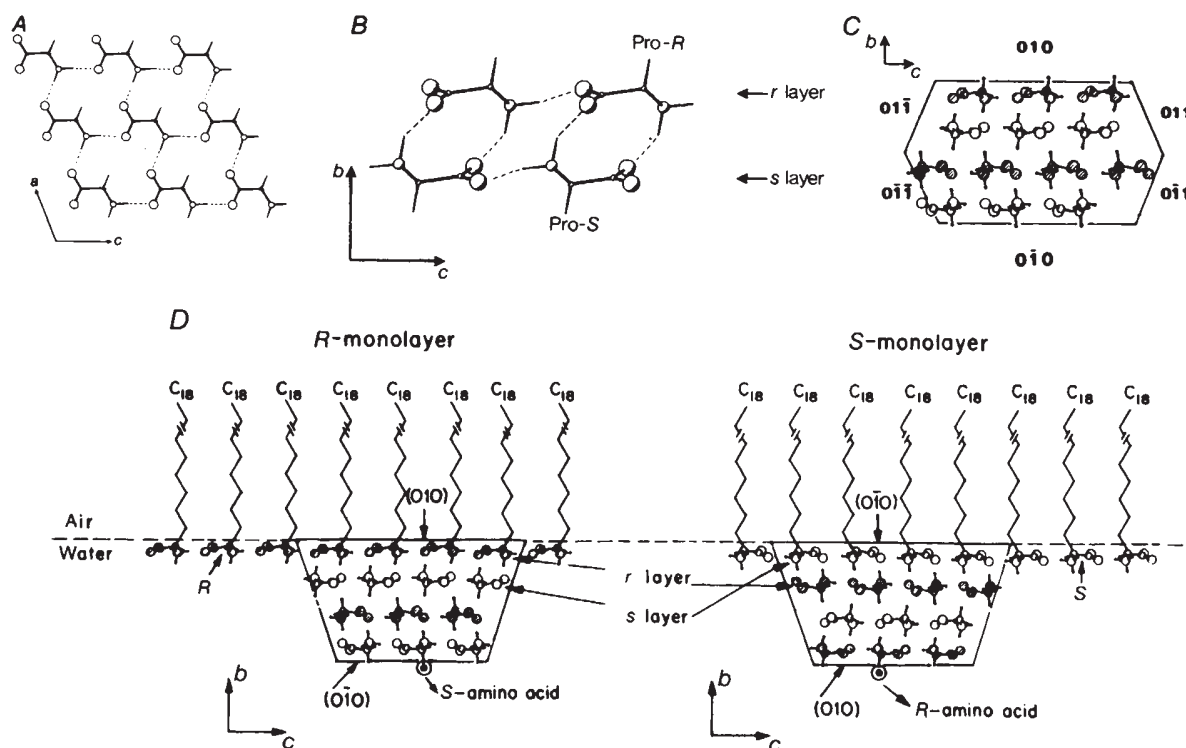


Fig. 1 A, An *ac* layer of hydrogen-bonded glycine molecules viewed along the *b*-axis. The layer is defined as *r* since the H atoms of the C—H bonds which emerge from the *ac* plane are pro-*R*. B, A centrosymmetric bilayer of hydrogen-bonded glycine molecules viewed perpendicular to the *bc* plane. The upper *r* layer is that shown in A. C, Packing arrangement of α -glycine, delineated by its crystal faces. The bilayers (B) are related by 2-fold screw symmetry along the *b*-axis. D, Schematic views of pyramidal crystals of glycine grown under compressed *R* and *S* amino-acid monolayers. The (010) face of glycine is attached to the *R* monolayer and the (0 $\bar{1}$ 0) face to the *S* monolayer. *S* and *R* amino acids are shown adsorbed from solution at the (0 $\bar{1}$ 0) and (010) faces respectively.

induce attachment of both (010) and (0 $\bar{1}$ 0) faces of glycine. These results on oriented crystal growth provide a new route for efficient amplification of optical activity of amino acids present in solution, through the enantioselective occlusion into the growing crystals of glycine¹ at water–air interfaces covered by a monolayer.

Glycine crystallizes from aqueous solution in the α -form (space group $P2_1/n$), exhibiting a bipyramidal habit¹. The glycine molecules form hydrogen-bonded chiral layers parallel to the *ac* plane² (Fig. 1A). These glycine layers (designated as *r* and *s*) are juxtaposed on one side of the *b* axis by centres of inversion about which the molecules are interlinked by N—H...O bonds into dimers to form centrosymmetric hydrogen-bonded bilayers (Fig. 1B). These bilayers are related by 2-fold screw symmetry to complete the crystal packing (Fig. 1C).

We anticipated that a densely packed Langmuir monolayer of amphiphilic homochiral α -amino acid molecules would exhibit in the aqueous subphase a hydrogen-bonded layer arrangement very similar to that in glycine, provided the hydrophobic moieties of the monolayer allow the neighbouring glycol head groups to be interlinked by N—H...O bonds. Consequently, a monolayer of, say, resolved *R*- α -amino acids packed as tightly as in an *ac* layer of glycine, should simulate an *r* layer of glycine molecules exposed at the (010) face of the crystal (see Fig. 1A, D), and thus might induce nucleation of a glycine crystal with its (010) face attached to the monolayer (Fig. 1D). By symmetry, the corresponding monolayer of *S*-amino acids should, in identical conditions, form an *s* layer and induce nucleation of a glycine crystal with its (0 $\bar{1}$ 0) face attached to the monolayer (Fig. 1D). Amphiphilic α -amino acids 1–5 (Table) were synthesized to test this hypothesis.

Condensed monolayers of compounds 1–3, measured over pure water, exhibited limiting areas per molecule in the range of 25–29 Å² (Fig. 2). These values straddle the molecular unit surface area (that is, $ac \sin \beta$) of 25.6 Å² of an *ac* layer of glycine molecules in its crystal structure. The compression isotherms of

the monolayers over aqueous solutions of glycine exhibited a distinct expanded character (Fig. 2), the effect being more pronounced with increasing concentration of glycine. This observation indicates the presence of interactions between the monolayer molecules and glycine from the subphase^{3,4}. At high compressions, all isotherms of a given compound converge (Fig. 2), indicating possible similarity in the final compressed structures of the respective films.

Monolayers 1–3 promoted immediate nucleation (within seconds after formation of the film) of pyramidal glycine crystals with their basal *ac* faces attached to the monolayer. When the *R*- α -amino acids were used, most pyramids were attached through their basal (010) faces to the monolayer and, by symmetry, when *S*- α -amino acids were used, enantiomorphous pyramids were formed with their basal (0 $\bar{1}$ 0) faces attached to the monolayer. Monolayers of racemic composition yielded attached pyramidal crystals of both types (Table 1).

With compound 5, which contains the cholestanoyl moiety, the limiting area per molecule in the monolayer is determined by the bulky steroid skeleton. It was measured to be 38 Å² (Fig. 2), a distinctly larger value than that of the glycol head group, thus precluding hydrogen bonding between the latter groups. Indeed, crystallization of glycine under such a monolayer was observed to occur at a much slower rate than with monolayers of compounds 1–3 (hours as opposed to seconds) and mostly in the bulk subphase rather than at the monolayer. In the few cases in which glycine also crystallized at the interface, these crystals exhibited both pyramidal and bipyramidal morphologies with no preferential orientation (Table 1).

For compound 4, containing a fluorocarbon chain, the situation is less clear. We might expect, with a limiting molecular surface area of the order of 30 Å² (Fig. 2)^{5,6}, that the glycol head groups would form a hydrogen-bonded layer, but with longer hydrogen-bonding distances than in the corresponding hydrocarbon monolayers. Monolayers of compound 4 of *S*

Table 1 Crystallization of α -glycine under compressed monolayers at the air-water interface

Monolayer	Limiting area per molecule ($\text{\AA}^2$)	Degree of orientation (%)	Face preferentially exposed to the monolayer	Rate of crystallization
1 $R\text{-CH}_3\text{-(CH}_2\text{)}_{15}\text{-CH-(NH}_3^+\text{)-CO}_2^-$ *	25	93	(010)	Fast (seconds)
$S\text{-CH}_3\text{-(CH}_2\text{)}_{15}\text{-CH-(NH}_3^+\text{)-CO}_2^-$ *	25	91	(0 $\bar{1}$ 0)	Fast
$R, S\text{-CH}_3\text{-(CH}_2\text{)}_{15}\text{-CH-(NH}_3^+\text{)-CO}_2^-$ *	25	(~50)	(010)	Fast
		(~50)	(0 $\bar{1}$ 0)	
2 $R\text{-CH}_3\text{-(CH}_2\text{)}_{17}\text{-OCO-CH}_2\text{-CH-(NH}_3^+\text{)-CO}_2^-$ †	29	>99	(010)	Fast
$S\text{-CH}_3\text{-(CH}_2\text{)}_{17}\text{-OCO-CH}_2\text{-CH-(NH}_3^+\text{)-CO}_2^-$ †	29	>99	(0 $\bar{1}$ 0)	Fast
3 $R\text{-CH}_3\text{-(CH}_2\text{)}_{17}\text{-OCO-(CH}_2\text{)}_2\text{-CH-(NH}_3^+\text{)-CO}_2^-$ †	29	>99	(010)	Fast
$S\text{-CH}_3\text{-(CH}_2\text{)}_{17}\text{-OCO-(CH}_2\text{)}_2\text{-CH-(NH}_3^+\text{)-CO}_2^-$ †	29	>99	(0 $\bar{1}$ 0)	Fast
4 $S\text{-CF}_3\text{-(CF}_2\text{)}_9\text{-(CH}_2\text{)}_2\text{-OCO-CH}_2\text{-CH-(NH}_3^+\text{)-CO}_2^-$ †	—‡	65-75	(0 $\bar{1}$ 0)	Fast
5 $S\text{-5-}\alpha\text{-Cholestan-3}\beta\text{-OCO-CH}_2\text{-CH-(NH}_3^+\text{)-CO}_2^-$ †	38	No orientation	—	Slow (hours)

The molecular unit surface area of α -glycine in an *ac* layer = $ac \sin \beta = 25.6 \text{ \AA}^2$. All crystallization experiments were performed at $20 \pm 2^\circ \text{C}$ by compressing the monolayers to their limiting areas per molecule over 4.66 M aqueous solutions of glycine. The monolayer area was kept constant during crystallizations. All monolayer-forming compounds have been unambiguously characterized by spectroscopic methods (NMR, MS, IR) and elemental analysis. All areas per molecule were measured over doubly distilled water at $20 \pm 2^\circ \text{C}$. The degree of orientation is defined as the percentage of crystals exposing a given attached face, that is, (010) or (0 $\bar{1}$ 0), to the monolayer. The rate of crystallization refers to the time elapsed between the compression of the film to its final area and the visual observation of first crystals attached to the film.

* Synthesized by aminolysis of α -bromostearic acid. The racemate was resolved into enantiomers by crystallization of the brucine salts of the *N*-formyl derivatives. The enantiomeric excess was determined to exceed 96% for monolayer 1(*R*) and 90% for monolayer 1(*S*) by $^1\text{H-NMR}$ and HPLC analyses of the diastereomeric *N*-formyl aminostearoyl *S*-phenylalanine methyl ester.

† Resolved compounds 2-5 were prepared by esterification of enantiomerically pure *N*-carbobenzoxy- α -benzyl aspartic or glutamic acids with the respective alcohols. The protective groups were cleaved by hydrogenolysis on Pd/C in 90% trifluoroacetic acid. The enantiomeric purity was determined to exceed 99% by gas chromatographic analyses¹³ of the respective *N*-trifluoroacetoxy diisopropyl aspartates or glutamates obtained by acidic hydrolysis of compounds 2-5 followed by derivatization.

‡ The area per molecule could not be reproduced because of technical problems (see Fig. 2).

configuration induced immediate crystallization of glycine. However, pyramids of both orientations were formed at the interface, that is, exhibiting (010) and (0 $\bar{1}$ 0) basal faces, yet with distinct preference for the latter (Table 1).

Crystallization experiments conducted with *R* monolayers of compound 2 in a range of areas per molecule showed complete orientation of glycine crystals at the monolayer up to surface areas five times its limiting value. This result suggests the formation of aggregates of very similar structure to the compressed monolayer which act as nucleation sites. The orientational effect is gradually lost at larger surface areas. Similar experiments conducted with monolayers of compounds 4 and 5 in a range of areas per molecule (up to five times their limiting areas) showed no orientational effect. These observations seem to rule out a crystallization mechanism involving a trivial surface concentration effect.

The three distinct modes of crystallization under the monolayers, fast oriented, fast unoriented and no crystallization or slow unoriented, imply a very sensitive dependence of the crystallization process on small changes in the structure of the monolayer. The fast oriented crystallization suggests the presence of a close fit between the arrangement of the glycol head groups in the monolayer and that in the *ac* layer of the glycine crystal. Departures from such a fit, possibly in only one direction, result in a significant loss of crystal orientation, without affecting the rate of crystallization. The need for cooperativity between the glycol moieties is clearly demonstrated in the third mode of crystallization, in which the nucleation rate is strongly inhibited. The observation of the fast crystallization of glycine with both (010) and (0 $\bar{1}$ 0) faces under a racemic monolayer may be interpreted not only in terms of a spontaneous enantiomeric segregation of the monolayer molecules into chiral domains⁷, similar in structure to that of the resolved enantiomers, but also in terms of other packing motifs arising from the miscibility of the *R* and *S* molecules. A detailed understanding of the structural correlation between the monolayers and the growing crystals must await atom-atom potential energy calculations and two-dimensional X-ray diffraction experiments using total external reflection of synchrotron radiation^{8,9}.

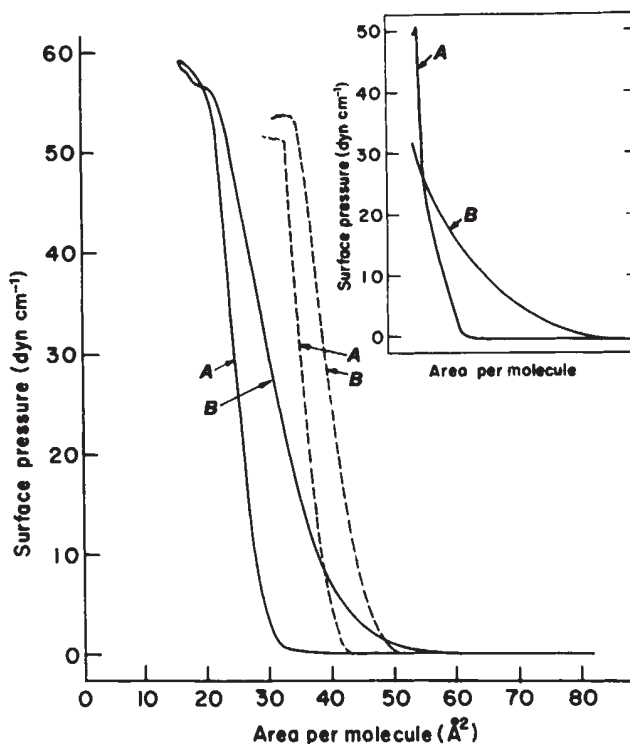
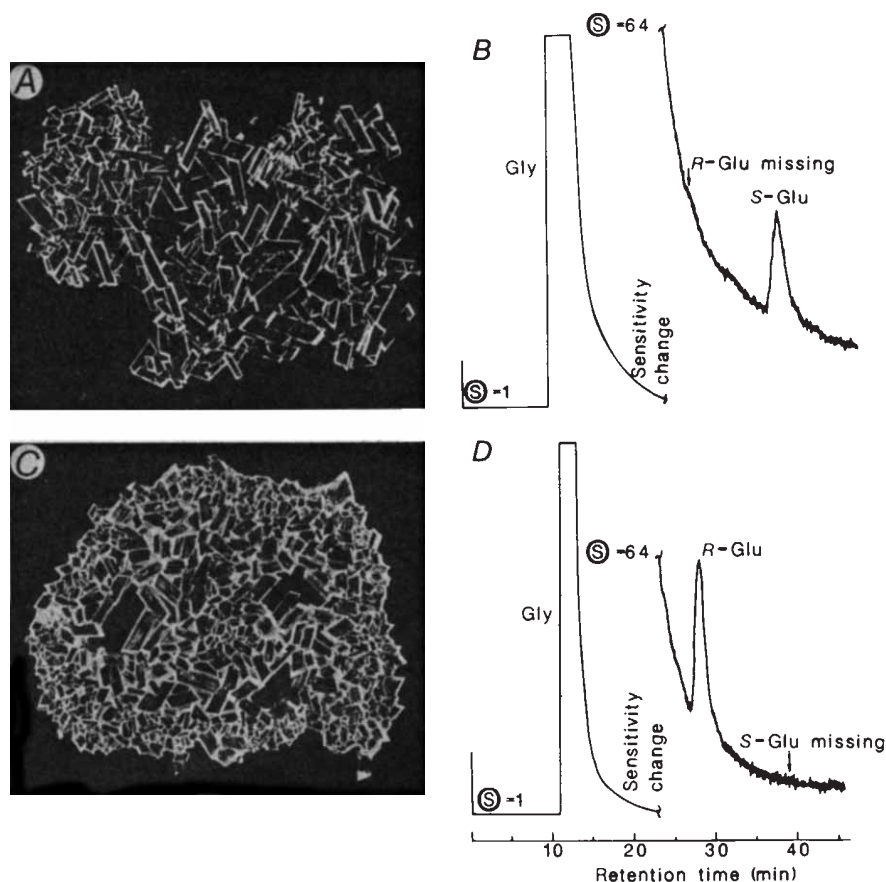


Fig. 2 Surface pressure-area isotherms measured with a thermostated circular trough equipped with a Wilhelmy balance (Mayer Feintech, Gottingen)¹⁴ at $20 \pm 2^\circ \text{C}$ on pure water (A) and a 2.66 M aqueous solution of glycine (B). —, Compound 2(*S*); ---, compound 5. The monolayers were spread from solutions in hexane (96%) and trifluoroacetic acid (4%). The insert displays isotherms of the fluorocarbon compound 4. The area per molecule of compound 4 could not be reproduced because of the formation of inhomogeneous spreading solutions. Expansion effects of monolayer films over salt solutions, similar to those observed here (curves B), have been reported for several systems^{3,4}.

Fig. 3 **A**, Crystals of glycine grown under a compressed monolayer of compound 2(*R*) from a 4.66 M aqueous solution of glycine containing 1% w/w *R*, *S*-Glu. All the crystals expose their (010) faces to the monolayer and exhibit the same enantiomorphous morphology. **B**, HPLC analysis of an ensemble of crystals grown as in **A** after washing with water. Conditions used: reverse-phase column (25 cm × 4.5 mm) self-packed with 5-μm Nucleosil C₁₈ (Machery Nagel). Mobile phase composition: aqueous solution of cupric acetate (4 mM) and *N,N*-dipropyl-L-Ala (8 mM) at pH 5.3–5.5 (ref. 15); post-column derivatization with *o*-phthalaldehyde and mercaptoethanol and fluorescence detection; flow rate, 0.4 cm³ min⁻¹, sensitivities $\odot$ as indicated¹⁵. **C**, Crystals of glycine grown under a compressed monolayer of compound 2(*S*) from a 4.66 M aqueous solution of glycine containing 1% w/w *R*, *S*-Glu. All the crystals expose their (0 $\bar{1}$ 0) faces to the monolayer and exhibit the opposite enantiomorphous morphology of those in **A**. **D**, HPLC analysis of an ensemble of crystals grown as in **C** after washing with water. Conditions as in **B**.



The ability to transfer structural information from monolayers to growing three-dimensional crystals suggests a new process for the amplification of optical activity at interfaces. It has been demonstrated previously that α -amino acids from a racemic mixture are enantioselectively occluded through the opposite {010} faces of a growing glycine crystal, leading to segregation along the *b*-axis^{1,10}. The *R*- α -amino acids are occluded only through the (010) face, whereas the *S*- α -amino acids are occluded only through the enantiotopic (0 $\bar{1}$ 0) face. Thus, when crystals of glycine were grown under an *R* monolayer of compound 2 in a solution containing racemic glutamic acid, only the *S* enantiomer was occluded into the growing glycine crystals attached to the monolayer (Fig. 3A, B), thus enriching the solution with *R*-glutamic acid. The analogous experiment performed with a monolayer of *S* configuration yielded only *R*-glutamic acid inside the glycine crystals (Fig. 3C, D), thus enriching the solution with the *S* enantiomer. The amplification effect is directly related to the number of molecular layers of the growing crystal.

The present studies are being extended to the orientational growth of other organic and inorganic crystals, also with the assistance of solid supported self-assembled^{11,12} and Langmuir-Blodgett films. This work will be reported elsewhere.

We thank Drs L. Addadi, Z. Berkovitch-Yellin, K. Hill and R. Popovitz-Biro, and Mrs E. Gati and I. Weissbuch, for stimulating discussions, and the Petroleum Research Fund administered by the American Chemical Society and the US/Israel Binational Science Foundation, Jerusalem, for financial support.

1. Weissbuch, I. *et al.* *Nature* **310**, 161–164 (1984).
2. Legros, J. P. & Kvik, A. *Acta crystallogr.* **B36**, 3052–3059 (1980).
3. Arsentiev, V. A. & Leja, J. in *Colloid and Interface Science* Vol. 5 (ed. Kerker, M.) 251–270 (Academic, New York, 1976).
4. Barnes, G. T. in *Colloid Science* Vol. 2 (ed. Everett, D. H.) 173–190 (Chemical Society, London, 1975).
5. Arrington, C. H. Jr. & Patterson, G. D. *J. phys. Chem.* **57**, 247–250 (1953).
6. Burnett, M. K. & Zisman, W. A. *J. phys. Chem.* **67**, 1534–1540 (1963).
7. Stewart, M. V. & Arnett, E. M. *Topics Stereochem.* **13**, 195–262 (1982).
8. Seul, M., Eisenberger, P. & McConnell, H. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5795–5797 (1983).
9. Als-Nielsen, J. & Pershan, P. S. *Nucl. Instrum. Meth.* **208**, 545–548 (1983).
10. Weissbuch, I. *et al.* *J. Am. chem. Soc.* **105**, 6615–6621 (1983).
11. Maoz, R. & Sagiv, J. *J. Colloid Interface Sci.* **100**, 465–496 (1984).
12. Netzer, L., Iscovici, R. & Sagiv, J. *Thin Solid Films* **99**, 235–241 (1983); **100**, 67–76 (1983).
13. Koenig, W. A., Benecke, I. & Sievers, S. *J. Chromat.* **217**, 71–79 (1981).
14. Fromherz, P. *Rev. scient. Instrum.* **46**, 1380–1385 (1975).
15. Weinstein, S. *Angew. Chem.* **21**, 218 (1982).
16. Addadi, L. *et al.* *Angew. Chem. int. Ed.* **24**, 466–485 (1985).

Regime and trace-element evolution of open magma chambers

Francis Albarede

Centre de Recherches Pétrographiques et Géochimiques,
Ecole Nationale Supérieure de Géologie, BP 20,
54501 Vandœuvre Cedex, France

O'Hara and co-workers^{1,2} have recently summarized some of the evidence accumulated for open-system, cyclic magmatic systems and provided useful geochemical equations for the replenishment/eruption cycles of an ideal magma reservoir. I show here that O'Hara's theory fits the situation where the volcanic eruption taps the more differentiated magma before the mixing stage. An alternative ideal model is discussed, which is applicable to the more common process of hybrid magma eruption. In mid-ocean ridge basalts, this new model may explain why large variations in incompatible elements coexist with moderate depletion of compatible elements.

Trace-element behaviour during fractional crystallization of a closed-system magma chamber is governed by the Rayleigh equation. Assuming an invariant mineralogical composition of

Received 12 August; accepted 3 September 1985.

1. Weissbuch, I. *et al.* *Nature* **310**, 161–164 (1984).
2. Legros, J. P. & Kvik, A. *Acta crystallogr.* **B36**, 3052–3059 (1980).
3. Arsentiev, V. A. & Leja, J. in *Colloid and Interface Science* Vol. 5 (ed. Kerker, M.) 251–270 (Academic, New York, 1976).
4. Barnes, G. T. in *Colloid Science* Vol. 2 (ed. Everett, D. H.) 173–190 (Chemical Society, London, 1975).
5. Arrington, C. H. Jr. & Patterson, G. D. *J. phys. Chem.* **57**, 247–250 (1953).

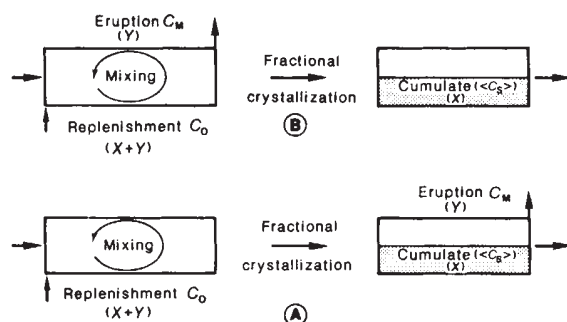


Fig. 1 The unit magmatic cycle linking replenishment, mixing, fractional crystallization and eruption for models A and B. X and Y represent the fraction crystallized and erupted in each cycle respectively. Model A is analogous to that presented by O'Hara¹ (equation (4) of the present work).

the cumulate, the concentration C_L of an element in the residual liquid is given by

$$C_L = C_{L0} (1 - X)^{D-1} \quad (1)$$

where X refers to the mass fraction of cumulate extracted from a parent magma with concentration C_{L0} , and D is the bulk solid/liquid partition coefficient. By mass balance, the concentration $\langle C_S \rangle$ in the bulk cumulate is given by:

$$\langle C_S \rangle = C_{L0} \frac{1 - (1 - X)^D}{X} \quad (2)$$

O'Hara's theory describes an ideal open magma reservoir as continuously differentiating, while being periodically tapped and refilled with fresh magma which mixes homogeneously with the remaining liquid. We define C_0 as the concentration in the fresh magma, C_B as the concentration in the erupted magma, X as the fraction removed as cumulate in each cycle, whilst Y is the fraction erupted as magma during the same period. The reference is the mass of liquid which fills the chamber when replenishment resumes. At steady state, equating the elemental fluxes results in

$$X \langle C_S \rangle + Y C_B = (X + Y) C_0 \quad (3)$$

Two extreme cases will be considered that differ in their timing of eruption, replenishment and differentiation (Fig. 1):

Case A: Step 1—the reservoir is filled with a batch of fresh magma which mixes homogeneously with residual liquids from previous cycles after expulsion of some of the previous residual liquid; step 2—closed system fractional crystallization takes place within the reservoir; step 3—part of the residual liquid is expelled and the system returns to step 1.

With the present notation, $C_B = C_L$. Using equations (1) and (2), equation (3) becomes

$$\frac{C_B}{C_0} = \frac{(X + Y)(1 - X)^{D-1}}{1 - (1 - X - Y)(1 - X)^{D-1}} \quad (4)$$

which is the same equation derived previously^{1,2}.

Case B: Step 1—the reservoir is filled with a batch of fresh magma which mixes homogeneously with residual liquids from previous cycles (as in case A); step 2—part of the mixed magma is immediately erupted; step 3—closed system fractional crystallization takes place within the reservoir and the system returns to step 1.

The model differs from that derived by O'Hara in the composition C_B of the erupted magma which here equals C_{L0} instead of C_L . Hence, equation (3) becomes

$$\frac{C_B}{C_0} = \frac{(X + Y)}{1 + Y - (1 - X)^D} \quad (5)$$

The equations for the steady-state with no crystallization ($X = 0$) are identical in both cases, whilst, as expected, they differ by a $(1 - X)^D$ term for the no-eruption limit ($Y = 0$).

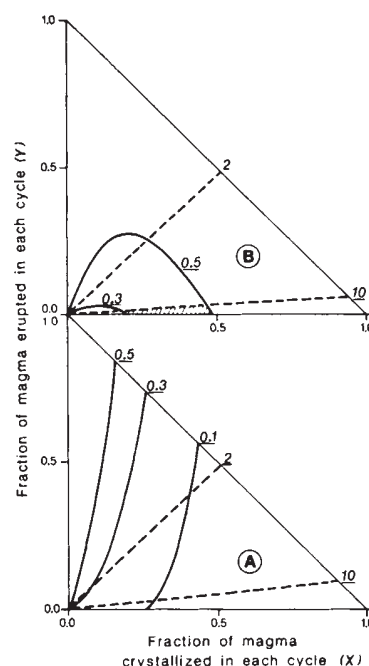


Fig. 2 Functional dependence of Y and X for a couple of elements having a bulk solid/liquid partition coefficient of $D=5$ (solid line) and $D=0.02$ (dashed line). Numbers underlined refer to different enrichment/depletion factors, that is, different C_B/C_0 ratios (see Fig. 1 and text for symbol explanation). The hatched area along the abscissa represents conditions of crystallization which result in both moderate depletion of transition elements and large enrichment of incompatible elements.

By analogy with the concept of residence time used in the marine environment, the characteristic number of cycles required for an element to 'approach' the cyclic concentration—say to within a few per cent—may be calculated as the ratio between the amount of this element initially present in the reservoir by the increment added in each cycle. This value, which is simply $C_B/(X + Y)C_0$ can be easily calculated from equations (4) and (5).

Both cases have a simple geological interpretation. In case A, the replenishment may be first considered as a consequence of the pressure release after the eruption of the previous cycle (Fig. 1). This sort of interpretation conflicts with geological evidence from volcanoes such as Kilauea³ or Krafla⁴: eruption is commonly preceded by a period of inflation of the chamber, which is likely to be caused by inflow of magma from below. Alternatively, eruption of a liquid may be triggered by the next pulse of fresh magma. In case A, eruption of differentiated melts implies that, for some reason (such as density contrast⁵), their mixing with the new batch of fresh magma takes place very slowly; the latter magma ponds at the base of the chamber and is involved in the eruption from the next cycle only. When mixing of high-density picritic liquids is achieved before closed system crystallization, which may not be systematic according to ref. 5, their trace-element evolution will be handled appropriately by O'Hara's equation (model A). In case B, mixing and subsequent eruption of hybrid magmas take place at the onset of the magmatic differentiation. For basaltic compositions, Sparks and co-workers⁵ predict that in many cases a turbulent mixing of fresh and differentiated magmas by the rise of plumes or diapirs occurs shortly before eruption, hence supporting model B.

Each model predicts an almost identical behaviour for incompatible elements but is extremely different for elements that are preferentially incorporated in the solid. Let us consider an element with a bulk solid/liquid partition coefficient $D=0.02$ and an element with $D=5$. Given an enrichment (or depletion) factor C_B/C_0 of the erupted magma relative to the fresh magma,

equations (4) and (5) may be solved for Y as a function of X . This function is plotted in Fig. 2 for cases A and B, and, for each element, assuming different enrichment (depletion) factors. As pointed out by O'Hara and Mathews², O'Hara's model (case A) cannot produce melts that have large enrichment in incompatible elements and moderate depletion of compatible elements unless both the melt fractions that crystallized and erupted in each cycle are fairly small. From field and petrological evidence, O'Hara suggests values in the range of 10^{-3} – 10^{-4} for Y and, from arguments which could also be deduced from a simple examination of my Fig. 2, concludes that X is also likely to be small.

Case B is expected to produce different trace-element patterns: inasmuch as the erupted liquid contains a significant fraction of undifferentiated fresh magma, depletion of compatible elements will not be as extreme as in case A. This model predicts the existence, along the X axis of Fig. 2, of a domain which reconciles wide variations of incompatible elements and moderate depletion of compatible elements associated with sizeable fractions of accumulated solids.

This conclusion may be illustrated using the genesis of mid-ocean ridge basalts (MORB). Although the nature of the MORB parent magma is still the subject of debate, there is some agreement that MORBs are not primitive melts which have issued from the mantle^{6,7}. Several authors have suggested that mixing fresh and residual magmas in reservoirs situated under the ridge system is an efficient mechanism for buffering the MORB composition. This mechanism would lead to characteristics of a moderately differentiated magma yet explain why primitive picritic basalts are so rare^{1,8}. Other authors have presented petrographical evidence of magma mixing and crystal/melt disequilibrium^{9–11}. Bearing in mind that some of the following assumptions are controversial, let us imagine a magma chamber which emits basalts with a Ni content of 150 parts per million (p.p.m.), typical of MORB basalts, while being periodically fed by a fresh—but not necessarily primitive—magma with 300 to 450 p.p.m. of Ni. Such fresh magma resembles the most 'primitive' basalts described on the ridges as glass of phenocryst inclusions^{10,12}. According to the present models and using the same definition as above, the Ni depletion factor would lie between 0.3 and 0.5 whilst a value of 5 would seem to be a reasonable estimate of its bulk solid/liquid partition coefficient D (ref. 13). O'Hara¹ and Bryan *et al.*¹⁴ have attributed wide variations of incompatible element concentrations in MORBs, reproducing certain features of plume-type basalts, to crystallization in an open chamber periodically replenished by new parental liquid. If an increase in the concentration of an incompatible element with $D = 0.02$ (Nd, for example) by a factor 10 is required, O'Hara's mechanism requires that both the erupted and the crystallized fractions be very small in each cycle (Fig. 2a), thereby implying that almost no chemical variation could be observed among samples produced at a given locality. This inferred homogeneity is a simple consequence of a nearly perfect steady state (both X and $Y \rightarrow 0$) and is clearly opposed to the variations observed in many sites over the ridge system, both on the kilometre scale¹⁵ and among melt inclusions in closely associated samples¹⁰. Deductions from the alternative model lead to different results (Fig. 2b): taking into account O'Hara's compelling evidence that Y must be small, a 10-fold enrichment of incompatible elements consistent with 10–30% of mineral fractionation and a moderate depletion of transition elements is now possible. A consequence of this model which has not yet been fully investigated is the larger number of cycles required for high- D elements to attain steady state.

I thank M. J. O'Hara and D. Velde for comments. This is CRPG contribution # 619.

Received 14 June; accepted 3 September 1985.

- O'Hara, M. J. *Nature* **266**, 503–507 (1977).
- O'Hara, M. J. & Mathews, R. E. *J. geol. Soc. Lond.* **138**, 237–277 (1981).
- Fiske, R. S. & Kinoshita, W. T. *Science* **165**, 341–349 (1969).
- Björnsen, A., Johnsen, G., Sigurdson, S., Thorbergsson, G. & Tryggvason, E. *J. geophys. Res.* **84**, 3029–3038 (1979).

- Sparks, R. S. J., Meyer, P. & Sigurdson, H. *Earth planet. Sci. Lett.* **46**, 419–430 (1980).
- Clarke, D. B. & O'Hara, M. J. *Earth planet. Sci. Lett.* **44**, 153–158 (1979).
- Hart, S. R. & Davis, K. E. *Earth planet. Sci. Lett.* **44**, 159–161 (1979).
- Sparks, R. S. J., Meyer, P. & Sigurdson, H. *Earth planet. Sci. Lett.* **46**, 419–430 (1980).
- Dungan, M. A. & Rhodes, J. M. *Contr. Miner. Petrol.* **67**, 417–431 (1978).
- Rhodes, J. M., Dungan, M. A., Blanchard, D. P. & Long, P. E. *Tectonophysics* **55**, 35–62 (1979).
- Walker, D., Shibata, T. & DeLong, S. E. *Tectonophysics* **70**, 111–125 (1979).
- Frey, F. A., Bryan, W. B. & Thompson, G. *J. geophys. Res.* **79**, 5507–5527 (1974).
- Hart, S. R. & Davis, K. E. *Earth planet. Sci. Lett.* **40**, 203–219 (1978).
- Bryan, W. B., Thompson, G. & Michael, P. J. *Tectonophysics* **55**, 63–85 (1979).
- Bryan, W. B. & Moore, J. G. *Bull. geol. Soc. Am.* **88**, 556–570 (1977).

Observations of laterally inhomogeneous anisotropy in the continental lithosphere

R. Kind

Seismologisches Zentralobservatorium, Krankenhausstr. 1–3, 8520 Erlangen, FRG

G. L. Kosarev, L. I. Makeyeva & L. P. Vinnik

Institute of Physics of the Earth, Academy of Sciences of the USSR, Moscow, USSR

Elastic anisotropy may be an important factor when seismic data are used to determine the structure, composition and dynamics of the Earth's interior^{1,2}. Azimuthal anisotropy has been found in the oceanic and, although less definitely, in the continental lithosphere from field experiments where the P_n -wave velocity was observed to vary with direction^{3–5}. These studies suggest a 180° periodicity in travel-time variations as the only criterion of the presence of azimuthal anisotropy. This criterion, however, may fail to distinguish between anisotropy and lateral inhomogeneity of the isotropic medium. A new method for measuring azimuthal anisotropy has been described recently⁶, based on the phenomenon of shear-wave splitting in anisotropic media, which suggests three criteria instead of one to identify azimuthal anisotropy. We have now applied this method to the records of permanent seismograph stations in southern Germany. Our results provide strong evidence for the presence of azimuthal anisotropy in the lithosphere of the region and reveal pronounced lateral variations in the parameters of anisotropy on a scale of about 200 km.

The method⁶ was published in Russian, so we will summarize it briefly. The idea is to analyse the transverse component T of converted shear waves like SKS or SmKS. These phases have travelled part of their path as P-waves and are therefore, in a laterally homogeneous and isotropic medium, strictly polarized as SV-waves. Should they have energy on the T component, this may be due to anisotropy or/and side refraction and scattering in a laterally inhomogeneous medium. We consider a transversely isotropic medium with a horizontal axis of symmetry, which forms a particular, relatively simple variety of azimuthal anisotropy. Such a medium is often used as a model in geophysical studies of azimuthal anisotropy. The wave fields set up in this model were examined using the Thomson–Haskell matrix method extended for anisotropic media^{7,8} and the following results were obtained.

If a plane SV-wave from an isotropic half-space passes from below into a plane horizontal, homogeneous, transversely isotropic layer with a horizontal axis of symmetry, then the incoming wave splits into two quasi-shear waves. Under realistic assumptions of the properties of the Earth's mantle (magnitude of anisotropy is of the order of several per cent, anisotropic layer thickness is of the order of 100 km, and the angle of incidence of the SV-wave is of the order of 10°), the time delay δt between the two quasi-shear waves at the top of the layer is of the order of a fraction of a second. For periods around 10 s, we may assume that $\omega \cdot \delta t \ll 1$, where ω is the circular frequency. Then, at the top of the layer, the quasi-shear waves interfere either constructively (in the radial component R) or destructively (in the transverse component T). The properties of the R

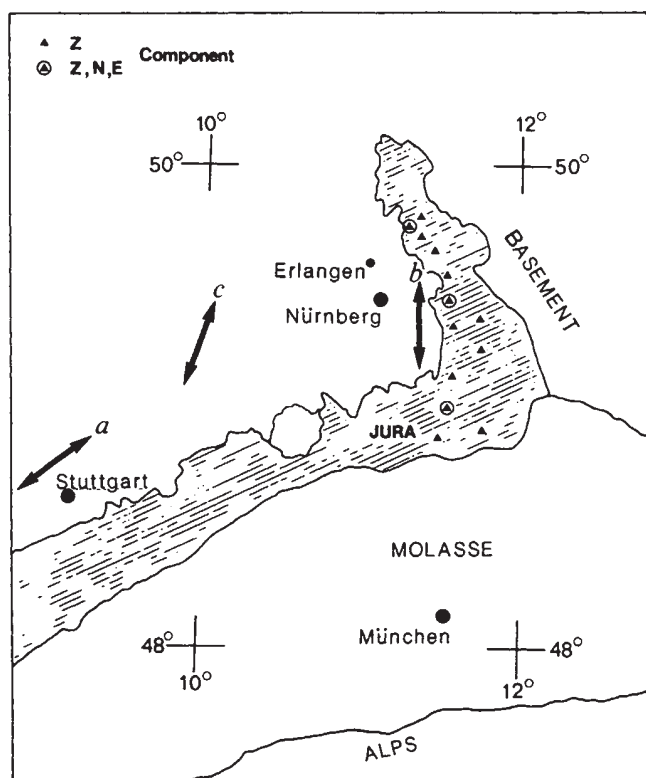


Fig. 1 Location map of the seismograph stations used and the major tectonic units of the area in southern Germany. The vertical components of the Gräfenberg array are marked by triangles, horizontal components by circles. A three-component station is located at Stuttgart. The arrows mark the directions of the fast velocity of the anisotropy observations. They are 50° at Stuttgart (a) and 0° at the Gräfenberg array (b). An averaged direction of 20° (c) results from a study of many P_n -arrival times over the entire area of southern Germany². Our study indicates that the general anisotropy of the area appears to break up into smaller units if a method with higher spatial resolution is used.

component are weakly dependent on the azimuth of propagation of the incoming wave, while major variations occur in the T component.

If the angle between the azimuth of the wave propagation and the axis of symmetry is $0, \pi/2, \pi$ or $3\pi/2$, the T component is equal to zero. If the angle is $\pi/4$ or $5\pi/4$ and the R component of the incoming wave is represented as $2 \cos(\omega t)$, then R and T at the top of the layer can be expressed in first approximation by

$$\begin{aligned} R(t) &\approx \cos(\omega t) + \cos(\omega t + \omega \delta t) \approx 2 \cos(\omega t) \\ T(t) &\approx \cos(\omega t) - \cos(\omega t + \omega \delta t) \approx (\omega \delta t) \sin(\omega t) \end{aligned}$$

These equations mean that the harmonic T component is shifted in time with respect to the R component by a quarter of a period, while the amplitude of the T component is proportional to frequency. The equations are also valid for the angles $3\pi/4$ and $7\pi/4$, but the expression for T changes sign. This change implies that the azimuthal variations of T are periodic with the period of π . The conclusions thus derived are in general corroborated by the calculation for other angles, while some minor complications resulting from sparse and noisy observational data may be neglected.

In summary, the long-period, harmonic T component of the displacement set up in the model considered is characterized in first approximation by the following properties⁶: (1) It is periodic as a function of the azimuth of the wave propagation; the dominant period is 180° . (2) It is shifted in time with respect to the radial component R by a quarter-period. (3) The T/R amplitude ratio is inversely proportional to the period.

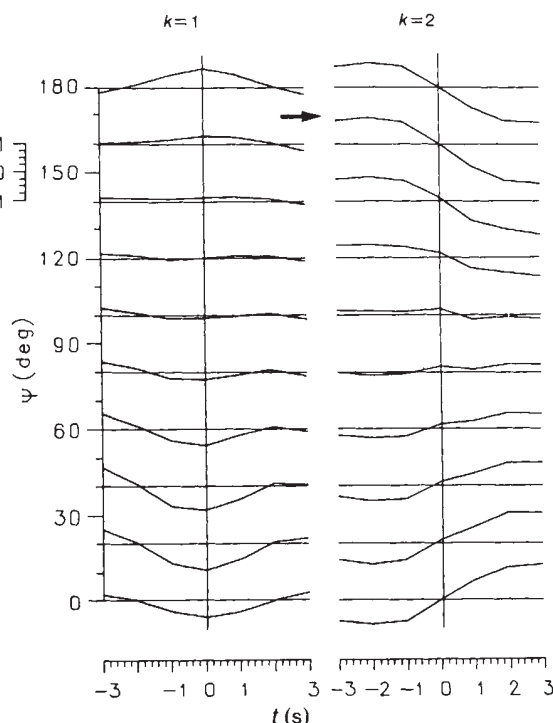


Fig. 2 Plots of $F(t, k, \psi)$ for the records of STU (modified from ref. 6). The largest deviations of $F(t)$ from 0 are observed in the right column ($k=2$) on the two top traces (marked by an arrow). These traces can be well approximated by the function $-0.12 \sin(2\pi t/10)$.

Thus, this approach suggests three independent diagnostic properties of azimuthal anisotropy instead of one in the conventional method. If all of them are observed, the probability is high that the observed effects are due to azimuthal anisotropy.

In order to extract the diagnostic properties of azimuthal anisotropy from seismic data, many earthquake recordings of the same seismograph station must be processed jointly. The combined processing of records of events with different magnitudes and different source signals requires standardization of the records. The normalized transverse component $\hat{T}(t)$ is computed for that purpose:

$$\hat{T}(t) = \int_{t_1}^{t_2} T(t+\tau) R(\tau) d\tau / \int_{t_1}^{t_2} (R(\tau))^2 d\tau$$

where the times t_1 and t_2 correspond to the beginning and end of the converted wave train and τ is the variable of integration. The use of $\hat{T}(t)$ instead of the transverse component $T(t)$ itself makes all the earthquakes directly comparable.

As an additional tool for the analysis of the azimuthal dependence, a harmonic analysis is performed

$$F(t, k, \psi) = \sum_i \hat{T}_i(t) \cos(k\varphi_i + \psi) / \sum_i (\cos(k\varphi_i + \psi))^2$$

where $\hat{T}_i$ corresponds to the i th seismic event, φ_i is the back azimuth of the epicentre of the i th event, k is the harmonic number and ψ varies in the range from 0 to 180° . The phase ψ_0 of the k th harmonic is given by the value of ψ , which corresponds to the maximum value of $|F(t, k, \psi)|$. Denoting by t_0 the value of t which corresponds to this maximum, we may define the amplitude of the k th harmonic as $|F(t_0, k, \psi_0)|$. Obviously the values of t_0 and ψ_0 are different for different harmonics. The minimum values of $|F(t, k, \psi)|$ correspond to $\psi = \psi_0 \pm \pi/2$ and can be used to locate ψ_0 .

$F(t, k, \psi)$ can now be examined for the three properties related to azimuthal anisotropy. The 180° periodicity of the T component is observed if $|F(t_0, 2, \psi_0)|$ is greater than $|F(t_0, 1, \psi_0)|$. The property of the phase shift between $R(t)$ and $T(t)$ is observed if the shape of $F(t, 2, \psi_0)$ is close to the sine function

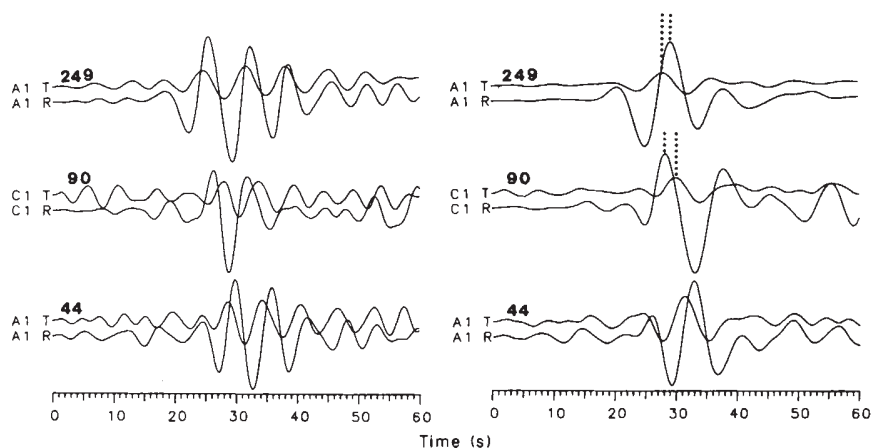
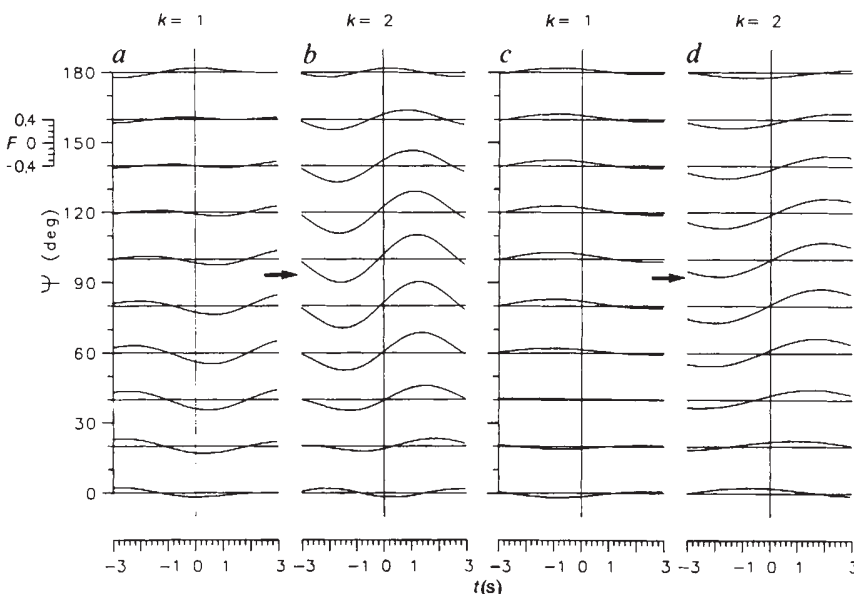


Fig. 3 SKS records of events 2, 15 and 8 of Table 1. The *R* and *T* components of these events are shown after passing through a shorter (left) and a longer (right) period bandpass filter. The numbers on the traces indicate back azimuths of the events in degrees. The letters *A* and *C* are the symbols for different sub-arrays. Note the phase shifts between the components (dotted lines).

Fig. 4 Plots of $F(t, k, \psi)$ for the records of GRF. A 4–8-s bandpass is used for *a* and *b*; and a 8–15-s bandpass for *c* and *d*. The largest deviations of $F(t)$ from 0 occur on the traces corresponding to $k=2$ and the values of ψ around 90° (marked by arrows). These traces can be well approximated by $0.4 \sin(2\pi t/6)$ (see *b*) and $0.3 \sin(2\pi t/8)$ (see *d*). The smallest values of $|F(t, 2, \psi)|$ correspond to the values of ψ around 0 and 180° conforming to $\psi_0 = 90^\circ$. Note the differences between the values of $F(t, 2, \psi)$ at GRF and STU (Fig. 2). They indicate pronounced differences in the parameters of azimuthal anisotropy.



$\sin(\omega t)$ or $-\sin(\omega t)$. The third property is observed if the values of $F(t_0, 2, \psi_0)$ are inversely proportional to the period. This requires, of course, that records in different frequency bands are available.

The direction of the fast velocity α (measured clockwise from north) can be found from the expression

$$\alpha = -\psi_0/2 + 3\pi/4$$

This expression was given in ref. 6 on the assumption that the shape of $F(t, 2, \psi_0)$ is close to $-\sin(\omega t)$. It may happen, however, that it is close to $\sin(\omega t)$. Then, instead of ψ_0 , one should use $\psi_0 + 180^\circ$.

The records of SKS and SKKS from the seismograph station STU (Stuttgart, Fig. 1) were available in the form of particle-motion plots for events of the years 1949–55 (ref. 9). It was impossible to filter these data, and the dominant periods were close to 10 s. The resulting plots of $F(t, k, \psi)$ (see Fig. 2) show that, for $k=2$, ψ_0 is close to 160 – 180° , $|F(t_0, 2, \psi_0)|$ is close to 0.12 and the diagnostic properties of azimuthal anisotropy are clearly observed: (1) The amplitude of the second harmonic $|F(t, 2, \psi_0)|$ is larger than $|F(t, 1, \psi_0)|$, implying that the influence of anisotropy is stronger than that of lateral inhomogeneity. (2) The shape of $F(t, 2, \psi_0)$ is close to the sine function, indicating the proper phase shift between the *R* and *T* components. The direction α is close to 50° . On the assumption of 7% anisotropy, the thickness of the anisotropic layer was estimated as 50 km (ref. 6).

The method was further applied to the digital broadband records of GRF (Gräfenberg array, Fig. 1; a description of this installation is given in ref. 10). The first step in processing the

records was to form beams of the radial and transverse components of the three-component stations of the array. The analysis was carried out in two frequency bands, one between 4 and 8 s and the other between 8 and 15 s. A list of the events used is given in Table 1.

Figure 3, showing examples of the filtered seismograms, indicates that considerable energy is present in the *T* component, which can only be caused by anisotropy or lateral inhomogeneity. Also clearly observable is the phase shift between the components, which can only be due to anisotropy. The properties of the plots of $F(t, k, \psi)$ (see Fig. 4) are in perfect agreement with those theoretically predicted for azimuthal anisotropy, namely: (1) the second harmonic ($k=2$) is much stronger than the first one; (2) the shape of the curve $F(t, 2, \psi_0)$ (ψ_0 is nearly 90°) is close to $\sin(\omega t)$; (3) the amplitudes of the second harmonic (0.4 for the shorter periods and 0.3 for the longer periods) are inversely proportional to the periods.

The observations at GRF provide decisive evidence for azimuthal anisotropy. The parameters of anisotropy, however, are different from those found at STU. The value of α at GRF can be estimated as 0° and the magnitude of the effect of anisotropy is twice as large as that found at STU. The difference between the values of α at GRF and STU is 50° , which is larger than the possible errors of measurement by an order of magnitude. This difference implies that the observed anisotropy is characteristic of the lithosphere and, perhaps, the top of the asthenosphere, since lateral variations of anisotropy on a scale of about 100 km in the deeper parts of the mantle would be averaged out by the summation procedure. We note that our estimates of α are close to those found from the teleseismic

Table 1 Events used in this study

Event	Date	Coordinates	Azimuth	Phase
1	25 Nov 76	19.5S 177.7W	17,197	SKKS, SKKS
2	30 Nov 76	20.6S 68.9W	249	SKS
3	21 Jul 77	53.8S 158.8E	114	PKS
4	22 Oct 77	28.0S 62.8W	240	SKS
5	17 Jul 78	15.0S 175.8W	12	SKKS
6	21 Aug 78	47.5S 32.3E	166	SKS
7	14 Mar 79	17.8N 101.3W	298	SKS
8	18 May 79	24.1N 142.4E	44	SKS
9	25 Jun 79	5.0S 145.6E	56	SKS
10	17 Oct 79	18.5N 145.4E	44	SKS
11	20 Oct 79	8.4S 115.8E	84	SKS
12	22 Oct 79	0.1N 126.0E	70	SKS
13	11 Dec 79	29.0N 141.0E	42	SKS
14	12 Dec 79	1.6N 79.3W	272	SKS
15	16 Apr 80	8.1S 108.8E	90	SKS
16	20 Jul 80	17.9S 178.6W	18	SKKS
17	11 Nov 80	51.5S 29.0E	169	SKS, SKKS
18	4 Sep 81	9.9N 124.0E	66	SKS
19	28 Sep 81	29.4S 179.0W	25,205	SKKS, SKKS
20	25 Oct 81	18.2N 102.0W	299	SKS
21	7 May 82	60.7S 20.8W	197	PS
22	11 Jul 83	60.9S 52.9W	211	SKS, SKKS
23	4 Oct 83	26.6S 70.8W	247	SKS
24	16 Nov 83	19.4N 155.5W	347	SKS, PS
25	22 Nov 83	0.4N 79.9W	271	SKS

Two phases were used for some events. Phases travelling in opposite directions across the globe have also been used in two cases.

P-wave travel-time residuals¹¹. The study of P_n -velocities on profiles crossing the same area in many directions has revealed a high velocity at the top of the mantle in an azimuth of 20° (ref. 5). This direction appears as the average of the substantially different directions at GRF and STU. The change of the direction of the high velocity between GRF and STU conforms to the shape of the strip of the outcrops of Jurassic rocks (Fig. 1). This conformity is either a coincidence or, more likely, the result of a relationship between the deep and shallow structure in the lithosphere.

The pronounced variations in the parameters of anisotropy within a small area imply that continental anisotropy is a relatively small-scale phenomenon compared with oceanic anisotropy. This difference is not surprising in view of the long and complicated history of the continental lithosphere. The term 'laterally inhomogeneous anisotropy' gives a proper description of the situation.

Finally, we note that the method⁶ has several advantages over conventional P_n -techniques. It does not require expensive field experiments and, as shown here, can provide higher lateral resolution. Moreover, it takes full advantage of the phenomenon of shear-wave splitting, which may be regarded as the most convincing evidence for the presence of anisotropy. Taking this into account, we recommend the method⁶ for widespread mapping of azimuthal anisotropy in the continental lithosphere. Data from dense, perhaps portable, broad-band seismograph networks would be ideally suited for this purpose.

This research was supported by the Deutsche Forschungsgemeinschaft. We thank M. Wyss and M. Hellweg for reading the manuscript.

Received 28 May; accepted 8 October 1985.

- Fuchs, K. *Phys. Earth planet. Inter.* **31**, 93-118 (1983).
- Crampin, S., Chesnokov, E. M. & Hipkin, R. G. *Geophys. J. R. astr. Soc.* **76**, 1-16 (1984).
- Raitt, R. W., Shor, G. G., Francis, T. J. G. & Morris, G. B. *J. geophys. Res.* **74**, 3095-3109 (1969).
- Shimamura, H. *Geophys. J. R. astr. Soc.* **76**, 253-260 (1984).
- Bamford, D. *Geophys. J. R. astr. Soc.* **49**, 29-48 (1977).
- Vinnik, L. P., Kosarev, G. L. & Makeyeva, L. I. *Proc. Acad. Sci. USSR* **278**, 1335-1339 (1984). (In Russian.)
- Crampin, S. *Geophys. J. R. astr. Soc.* **21**, 387-402 (1970).
- Kosarev, G. L., Makeyeva, L. I., Savarensky, Ye. F. & Chesnokov, Ye. M. *Izvestiya, Earth Phys.* **15**, 102-110 (1979) (AGU translation).
- Richter, G. *Über die Polarisation der SKS- und SKKS-Welle* (Habilitationsschrift, Halle, Wittenberg, 1957).
- Harjes, H. P. & Seidl, D. *J. geophys. Res.* **44**, 511-523 (1978).
- Dziewonski, A. M. & Anderson, D. L. *J. geophys. Res.* **88**, 3296-3314 (1983).

Near-synchronicity of New Zealand alpine glaciations and Northern Hemisphere continental glaciations during the past 750 kyr

C. S. Nelson, C. H. Hendy, G. R. Jarrett
& A. M. Cuthbertson

School of Science, University of Waikato, Hamilton, New Zealand

The Brunhes magnetochron (0-730 kyr) at Deep Sea Drilling Project Site 594 off eastern South Island, New Zealand, comprises up to ~100 m of alternating units of pelagic and hemipelagic ooze formed during interglacial and glacial periods respectively. The youngest hemipelagic interval corresponds to glacial isotope stage 2 and can be correlated with extensive glacial and glacio-fluvial deposits of the late Last Glaciation (~27-13 kyr BP) on South Island. Inferring the same genetic link for the other hemipelagic sediment units down core, we show here that as many as 12 major episodes of expanded alpine glaciation have occurred on South Island in the past 730 kyr. Each of these glacial episodes coincided with a period of significant ¹⁸O enrichment (+2‰) in the core resulting from changes in ocean isotope composition associated with major growth phases of the Northern Hemisphere ice sheets. The inferred history of alpine glaciations on this remote South Pacific island appears to be closely phase-locked (±3 kyr) to world-wide climate changes.

Site 594 of the Deep Sea Drilling Project Leg 90 is located on the southwestern margin of the Chatham Rise, 300 km east of New Zealand's South Island, in a water depth of 1,204 m (Fig. 1). The site lies in the sub-antarctic water mass immediately south of the Subtropical Convergence. The upper 170 m of the section was drilled with 95% recovery using an hydraulic piston corer¹ and provided largely undisturbed cores comprising conspicuously alternating units of light-coloured pelagic and dark-coloured hemipelagic oozes of Pliocene-Quaternary age^{2,3}. Seismic profiles⁴ demonstrate that this sedimentary sequence is

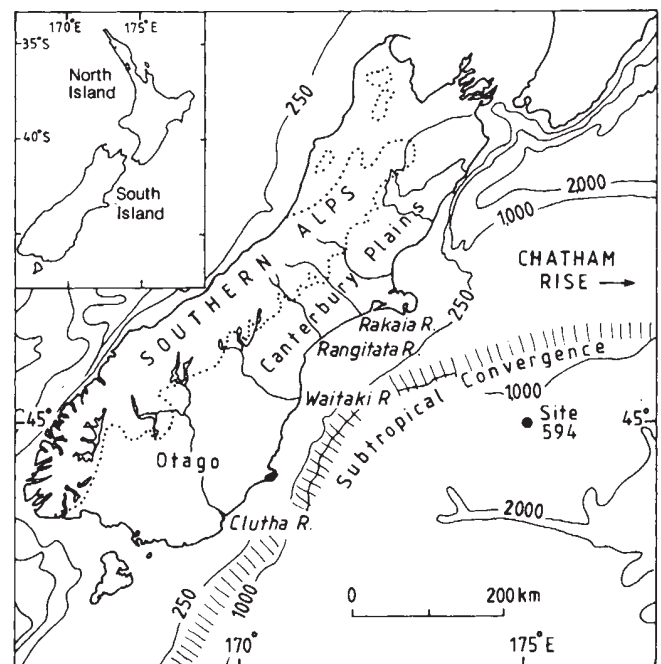


Fig. 1 Location map for DSDP Site 594 (45°31.41'S, 174°56.88'E) off South Island, New Zealand, showing also the approximate position of the oceanic Subtropical Convergence and some relevant physiographical features on South Island. The dotted line on land shows the probable extent of glacial ice at about 18 kyr BP (ref. 16).

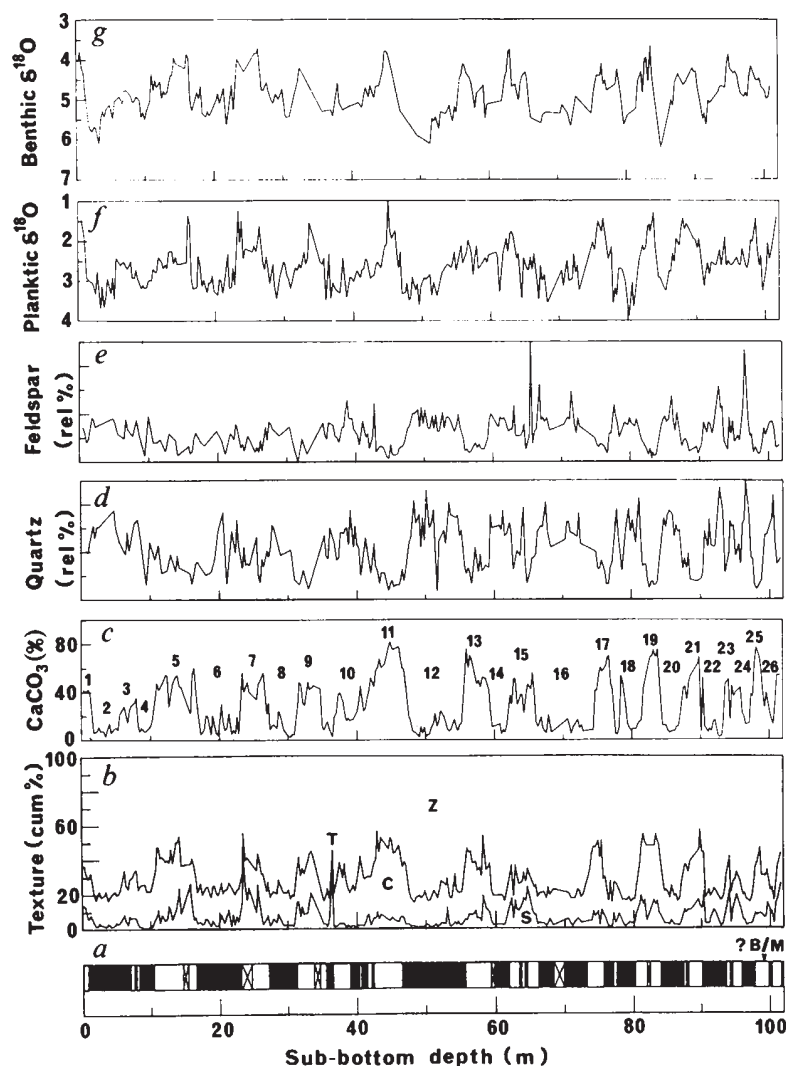


Fig. 2 Colour (a), texture (b), calcium carbonate (c), quartz (d), plagioclase feldspar (e) and planktonic and benthic foraminiferal oxygen-isotope (f and g respectively) records for the late Quaternary section of Site 594 to the Brunhes-Matuyama (B/M) boundary (730 kyr), tentatively placed at about 99.4 m sub-bottom depth³, although other options are possible. White- and black-coloured intervals (a) are dominated by pelagic and hemipelagic oozes respectively; crossed intervals represent missing core. The mean age resolution of sample points in most plots is ~2.4 kyr. In the cumulative percent textural plot (b): S, sand (>0.06 mm); C, clay (<0.004 mm); Z, silt (0.004–0.06 mm); T, tephra layer. Numbers 1–26 in the carbonate record (c) refer to major intervals of high (odd numbers) and low (even numbers) calcium carbonate content. Note that compared with the light-coloured pelagic ooze intervals the dark-coloured hemipelagic sediment units are characterized by depletion in sand and clay sizes, low carbonate contents, relatively increased amounts of quartz and feldspar, and significant ^{18}O enrichment, properties consistent with their formation during full glacial episodes. Analytical procedures and data are recorded in refs 5, 6.

regionally extensive off eastern central South Island. Here we discuss only the late Quaternary interval of core, which is complete and exceptionally thick, the Brunhes normal magneto-chron (0–730 kyr) possibly extending to about 99.4 m sub-bottom depth³.

The pelagic sediments are mainly impure foraminifer-bearing nanofossil oozes whose relative enrichment in both sand- and clay-sized material (Table 1) reflects their increased proportions of foraminiferal tests and coccolith plates respectively. The hemipelagic sediment units are more variable in composition, but are typically nanofossil-, diatom- and/or sponge spicule-bearing terrigenous clayey silts with a preponderance of silt-sized quartz, sodic plagioclase, chlorite and mica minerals (Table 1).

The bulk sediment texture, calcium carbonate content and relative abundance (from X-ray diffraction) of quartz and feldspar have been determined for 315 samples collected at 30 cm intervals through the section^{5,6}, equivalent to a mean age resolution of one sample per 2.4 kyr over the past 750 kyr. These sedimentary parameters exhibit strong in-phase cyclical fluctuations down core (Fig. 2a–e). Using the carbonate stratigraphy as a basis, 26 major sedimentary alternations occur in the top 100 m of core (Fig. 2c), 13 of predominantly pelagic character (odd numbers 1 to 25) and 13 with mainly hemipelagic properties (even numbers 2 to 26).

Oxygen-isotope measurements were made on both planktonic (*Globigerina bulloides*) and benthic (mainly *Uvigerina* sp.) foraminifers from the same samples used for the sedimentary analyses⁵ (Fig. 2f, g). Despite superimposed high-frequency

fluctuations, the isotope curves display overall in-phase cyclical variations of about 2‰ in ^{18}O enrichment and depletion which in turn match the sedimentary fluctuations corresponding to the hemipelagic and pelagic ooze units respectively. Such extreme variations in the isotope composition of the ocean, as recorded in the foraminiferal tests, are controlled primarily by changes in global ice volume^{7,8} and clearly associate the hemipelagic sediments with glacial conditions and the pelagic ones with interglacial periods.

This strong interrelationship between climate and sedimentation suggests that the abundant terrigenous material in the hemipelagic deposits resulted primarily from greatly increased rates of erosion on the South Island during glacial periods and the rapid transferral of this terrigenous detritus to off-shore areas where sedimentation rates increased and the carbonate content of bottom sediments was diluted (Table 1). Direct evidence for greatly increased rates of erosion and supply of sediment from South Island comes only from the nature and distribution of on-land^{9,10} and off-shore^{11,12} deposits associated with the last major glacial advances in New Zealand, between ~27 and 13 kyr BP (ref. 13). This timing matches closely isotope stage 2 (ref. 8), which is clearly represented by the hemipelagic deposit between ~1.5 and 5.5 m sub-bottom depth in our core (unit 2 in Fig. 2c), and is discernible also in other cores from the region¹¹. At this time the lowlands of South Island east of the Southern Alps accumulated thick deposits of greywacke and schistose gravels and sands^{14–16}, commonly in the form of extensive coalescing outwash fans, such as Canterbury Plains (Fig. 1). The major eastward-draining South Island rivers also

Table 1 General sediment properties of the pelagic and hemipelagic oozes in the late Quaternary section at DSDP Site 594

Property	Pelagic oozes	Hemipelagic oozes
Colour	Light blueish-grey	Dark greenish-grey
Bulk sediment grain size		
% Sand (>63 μm)	10-30	0-10
% Silt (4-63 μm)	50-70	70-90
% Clay (<4 μm)	25-30	10-25
Calcium carbonate (%)	40-80	0-25
Nannofossils (%)	Abundant (40-75)	Rare to common (1-25)
Foraminifers (%)	Common (5-25)	Rare (1-5)
Siliceous microfossils (%) [*]	Absent to rare (0-10)	Rare to common (5-30)
Terrigenous minerals (%) [†]	Rare to common (10-40)	Abundant (30-75)
Sedimentation rates (cm kyr^{-1}) ⁵	5-10	10-35

^{*} Includes sponge spicules, diatoms, radiolarians and silicoflagellates.

[†] Includes quartz, plagioclase, chlorite and micas.

generate high proportions of suspended load^{17,18}. For example, 94% of the quartzfeldspathic chlorite schist supplied originally as bedload to the Clutha River (Fig. 1), a major source of the marine sediment off Otago and south Canterbury^{12,19}, is abraded to silt and finer sizes during its course to the sea¹⁷. During glacial periods the reduced shelf width accompanying lowered sea levels undoubtedly enhanced deep-water mud deposition^{11,12}. Moreover, the occurrence of widespread deposits of loess of Last Glacial age in east Canterbury and Otago²⁰, and the significant increase in the content of aeolian quartz in 18-15-kyr-old deep-water sediments east of New Zealand compared with their modern counterparts^{21,22} suggest that a significant proportion of the terrigenous silt in the hemipelagic sediments at Site 594 may be of aeolian origin, picked up from the extensive outwash surfaces on eastern South Island and transported offshore by the prevailing westerly wind system, which was intensified during glacial times¹³.

Thus hemipelagic unit 2 at Site 594 is the off-shore sedimentary result of climatic deterioration associated with the latest phase of extended alpine glaciation in South Island, and is directly correlatable with isotope stage 2 reflecting the last period of major expansion of the Northern Hemisphere ice sheets. The similar nature of the hemipelagic sediment intervals throughout the section (Fig. 2) strongly suggests that every hemipelagic unit reflects a major period of increased erosion and outwash aggradation associated with expanded glaciation in South Island. Moreover, because each hemipelagic interval corresponds to a period of significant (+2‰) ¹⁸O enrichment (Fig. 2f, g) reflecting mainly changes in the isotope composition of the ocean due to an increase in the volume of global ice, then each episode of alpine glaciation coincided with a world-wide period of glacio-eustatically lowered sea level. The late Quaternary marine oxygen-isotope changes have been controlled primarily by the growth and retreat of continental ice sheets in the Northern Hemisphere^{7,23}; an accord between the marine ¹⁸O record and the northern European glacial record inferred from loess sequences has been demonstrated by Kukla²⁴.

That every major ice growth phase indicated by our oxygen-isotope data (Fig. 2f, g) can be closely matched through the hemipelagic units to a period of expanded alpine glaciation on a small temperate-latitude island in the Southern Hemisphere is evidence that the latter does not reflect simply local or isolated climatic events, but rather world-wide climate change. Moreover, since no consistently discernible phase difference is evident between the isotope and lithological records (Fig. 2), the climatic and sedimentary response in New Zealand to fluctuations in the world's continental ice sheets has been rapid, within the 2-3-kyr period represented by our sample spacing.

Finally, it is unfortunate that some uncertainty surrounds the exact position of the Brunhes-Matuyama boundary in this high-

resolution record²⁵, the location chosen here (Fig. 2a) being that adopted in Leg 90 of the *Initial Reports of the Deep Sea Drilling Project*^{3,5,25,26}. If this placement is correct, and the Brunhes sedimentary record at Site 594 is genuinely free of unconformities, then there have been 12 full glacial periods and 13 major interglacial episodes over the past 730 kyr (Fig. 2c), more than is implied by the standard 19 oxygen-isotope stage scheme for this time interval⁸. However, it might be argued from the shape and amplitudes of the isotope and carbonate records in Fig. 2 that the Brunhes-Matuyama boundary lies nearer 84 m sub-bottom (W. F. Ruddiman, personal communication, 1985), and that the first 19 sedimentary units defined here (Fig. 2c) match the recognized number of isotope stages for the Brunhes magnetochron. This important chronological aspect of the late Quaternary climate record at Site 594 remains to be resolved and is under investigation.

We thank Professor Michael Selby (University of Waikato) and Dr William Ruddiman (Lamont Observatory) for helpful comments, the New Zealand University Grants Committee for financial assistance (grant 88/125) of the project, and the Deep Sea Drilling Project for logistic support.

Received 21 June; accepted 10 September 1985.

1. Prell, W. L. *et al. Init. Rep. DSDP* **68**, 5-6 (1982).
2. Leg 90 Scientific Party *Geotimes* **28**, 16-19 (1983).
3. Kennett, J. P. *et al. Init. Rep. DSDP* **90**, (in the press).
4. Lewis, K. B., Bennett, D. J., Herzer, R. H. & von der Borch, C. *Init. Rep. DSDP* **90**, (in the press).
5. Nelson, C. S., Hendy, C. H., Cuthbertson, A. M. & Jarrett, G. R. *Init. Rep. DSDP* **90**, (in the press).
6. Jarrett, G. R. thesis, Univ. Waikato, Hamilton (1985).
7. Shackleton, N. J. & Opdyke, N. D. *Quat. Res.* **3**, 39-55 (1973).
8. Imbrie, J. *et al. in Milankovitch and Climate Change* (eds Imbrie, J. & Berger, A.) (Elsevier, Amsterdam, 1985).
9. Suggate, R. P., Stevens, G. R. & Te Punga, M. T. (eds) *The Geology of New Zealand* vol. 2 (NZ Geological Survey, DSIR, Wellington, 1978).
10. Burrows, C. J. *Mauri Ora* **6**, 69-96 (1978).
11. Griggs, G. B., Carter, L., Kennett, J. P. & Carter, R. V. *Bull. geol. Soc. Am.* **94**, 791-797 (1983).
12. Herzer, R. H. *N.Z. J. Geol. Geophys.* **22**, 391-406 (1979).
13. Salinger, M. J. in *Late Cainozoic Palaeoclimates of the Southern Hemisphere* (ed. Vogel, J. C.) (Balkema, Rotterdam, 1984).
14. Suggate, R. P. in *Landforms of New Zealand* (eds Soons, J. M. & Selby, M. J.) (Longman, Auckland, 1982).
15. Fitzharris, B. B., Mansergh, G. D. & Soons, J. M. in *Landforms of New Zealand* (eds Soons, J. M. & Selby, M. J.) (Longman, Auckland, 1982).
16. Fleming, C. A. *The Geological History of New Zealand and its Life* (University Press, Auckland, 1979).
17. Adams, J. *Bull. geol. Soc. Am.* **91**, Pt II, 1-144 (1980).
18. Gibb, J. G. & Adams, J. *N.Z. J. Geol. Geophys.* **25**, 335-352 (1982).
19. Andrews, P. B. *N.Z. J. Geol. Geophys.* **16**, 793-830 (1973).
20. McCraw, J. D. in *Quaternary Studies* (eds Suggate, R. P. & Cresswell, M. M.) *R. Soc. N.Z. Bull.* **13**, 35-44 (1979).
21. Thiede, J. *Geology* **7**, 259-262 (1979).
22. Stewart, R. B. & Neall, V. E. *Nature* **311**, 47-48 (1984).
23. Shackleton, N. J. & Opdyke, N. D. in *Investigation of Late Quaternary Paleoclimatology and Paleoclimatology* (eds Cline, R. M. & Hays, J. D.) *Geol. Soc. Am. Mem.* **145**, 449-464 (1976).
24. Kukla, G. J. *Earth-Sci. Rev.* **13**, 307-374 (1977).
25. Barton, C. E. & Bloemendal, J. *Init. Rep. DSDP* **90**, (in the press).
26. Ciesielski, P. F. *Init. Rep. DSDP* **90**, (in the press).

First Mesozoic mammal from Australia—an early Cretaceous monotreme

Michael Archer*, Timothy F. Flannery†, Alex Ritchie† & R. E. Molnar‡

* School of Zoology, University of New South Wales, PO Box 1, Kensington, New South Wales, Australia 2033

† Australian Museum, 6-8 College Street, Sydney, New South Wales, Australia 2000

‡ Queensland Museum, Fortitude Valley, Queensland, Australia 4006

Here we describe Australia's first known Mesozoic mammal and the first known early Cretaceous mammal from Gondwanaland. *Steropodon galmani* n. gen. and sp., discovered in early Cretaceous sediments at Lightning Ridge, New South Wales, Australia, appears to represent an ornithorhynchid-like monotreme. This

discovery represents the first record of a fossil mammal from Australia that is older than 22.4 ± 0.05 Myr^{1,2} and the specimen is, by more than 85 Myr, the oldest known monotreme. As the oldest monotreme, it will necessitate a radical revision of present understanding about dental homology in the middle Miocene *Obdurodon insignis*, the only fossil monotreme previously known to have had teeth³. The structure of *S. galmani* supports one current view⁴ that monotremes, one of three groups of living mammals (the other two being marsupials and placentals), are phylogenetically close to the other groups of living mammals.

The opal-bearing sediments at Lightning Ridge consist of a suite of siltstones and poorly sorted sandstones referred to the early Cretaceous Griman Creek Formation⁵. In New South Wales this formation includes the Wallangulla Sandstone Member⁵, the unit from which the new monotreme was obtained. The Griman Creek Formation is dated as middle Albian, based on palynological evidence⁹. This age is supported by the discovery in the formation of hypsilophodontid dinosaurs that are very similar to those found in the Strzelecki Group of Victoria, which has been fission-track-dated as Aptian-Albian⁷. At Lightning Ridge the Griman Creek Formation was apparently deposited in the shallow estuary of a westward-flowing river. Some of the formation's invertebrate fossils are clearly marine but others, including viviparid snails, indicate a freshwater influence. Cross-bedding is present, as are apparent impressions of plant roots^{5,8}. Many of the tetrapod bones indicate transport. In addition to *S. galmani*, the vertebrate fauna from Lightning Ridge consists of teleosts, ceratodontid lungfish (*Ceratodus wollestoni* and *Neoceratodus forsteri*), turtles, plesiosaurs, crocodilians ('*Crocodylus*' *selasphensis*), and theropod (*Rapator ornitholestoides*), ornithopod (*Fulgurotherium australe*) and sauropod dinosaurs^{6,10,11}. There is no evidence that any of the fossils represent an intrusion of later material into the Griman Creek Formation.

The dental terminology¹² used here for *S. galmani* assumes that the teeth are tribosphenic and the cusps are homologues of topographically analogous cusps in therian mammals.

Class Mammalia

Subclass Monotremata

Family ?Ornithorhynchidae

Genus and species: *Steropodon galmani* n. gen. and sp. (Figs 1, 2).

Etymology: *Sterope* is Greek for 'flash of lightning' and *odon* is Greek for 'tooth', the name being an allusion to Lightning Ridge, the type locality, as well as to the opalescent nature of the fossil teeth.

Holotype: Australian Museum Palaeontology collection no. AM F66763, a right dentary fragment with M₁₋₃ in place, obtained from an opal miner by David and Alan Galman in 1982 or 1983. This specimen is, like all fossils from this locality, a natural opal pseudomorph of the original. Accordingly, there is no preservation of the internal bone or tooth structure although cavities in the original, such as alveoli and the mandibular canal, have been differentially preserved. The teeth preserved in the holotype have been interpreted as M₁₋₃ because the alveolus anterior to the anteriormost molar is very small. This indicates that there was an abrupt change in shape between M₁ and the tooth presumed to be a premolar situated anterior to it.

Locality: Lightning Ridge, New South Wales, Australia.

Stratigraphy: Wallangulla Sandstone Member of the Griman Creek Formation.

Age: Albian⁹.

Diagnosis: This taxon differs from all other monotremes (species of *Obdurodon*, *Ornithorhynchus*, *Tachyglossus* and *Zaglossus*) in its retention of a cristid obliqua and talonid basin and in its possession of a pre-entocristid. It differs from non-monotrematous mammals in its combination of the following features: three lower molars, an antero-posteriorly very compressed trigonid, a transverse metacristid, a well-developed and wide talonid, a pre-entocristid/cristid obliqua/distal metacristid conjunction, absence of a paraconid on M₁, absence of

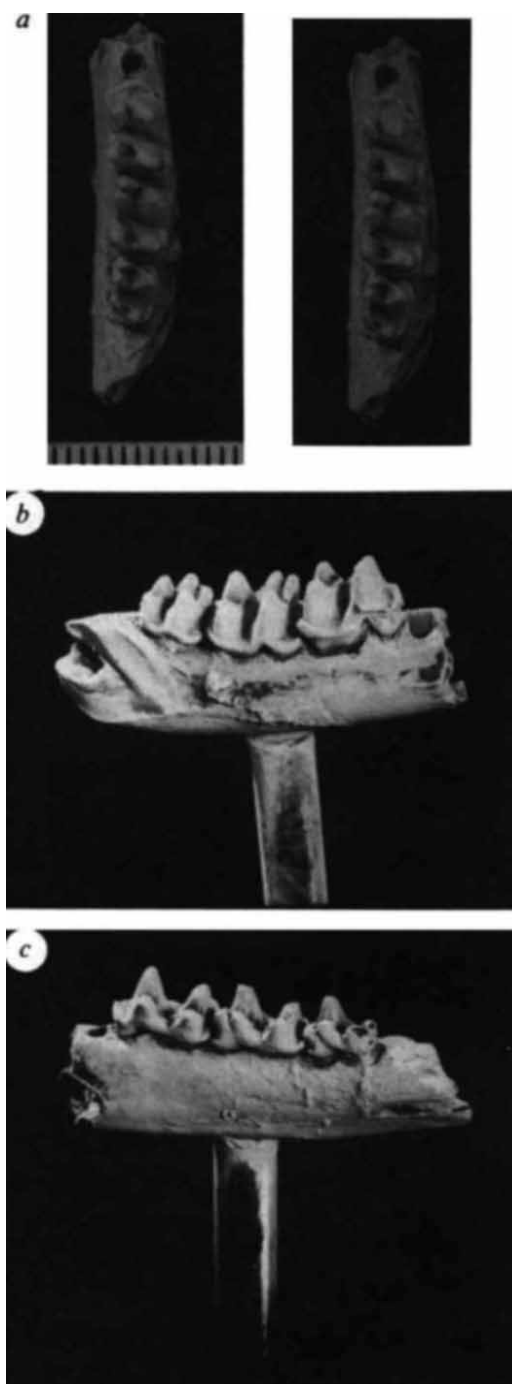


Fig. 1 The holotype of *Steropodon galmani* n. gen. and sp. in occlusal view (a), buccal view (b) and lingual view (c).

hypoconulids, presence of a distinctive distal metacristid and molars that exhibit essentially transverse shearing crests.

Description: While a more detailed description is in preparation, the above diagnosis, the comments made below and the illustrations (Figs 1, 2) will suffice for present purposes.

A posterior, elongate, shallow horizontal depression on the lingual side of the dentary extends from the posterior broken edge of the dentary forward to the level of the posterior root of M₃; this may represent a groove for the meckelian cartilage or it may be the result of pre-depositional breakage that seems to have affected other areas of the posterior part of the jaw fragment.

We conclude that *S. galmani* is a monotreme for the following reasons. Its antero-posteriorly very compressed trigonid,

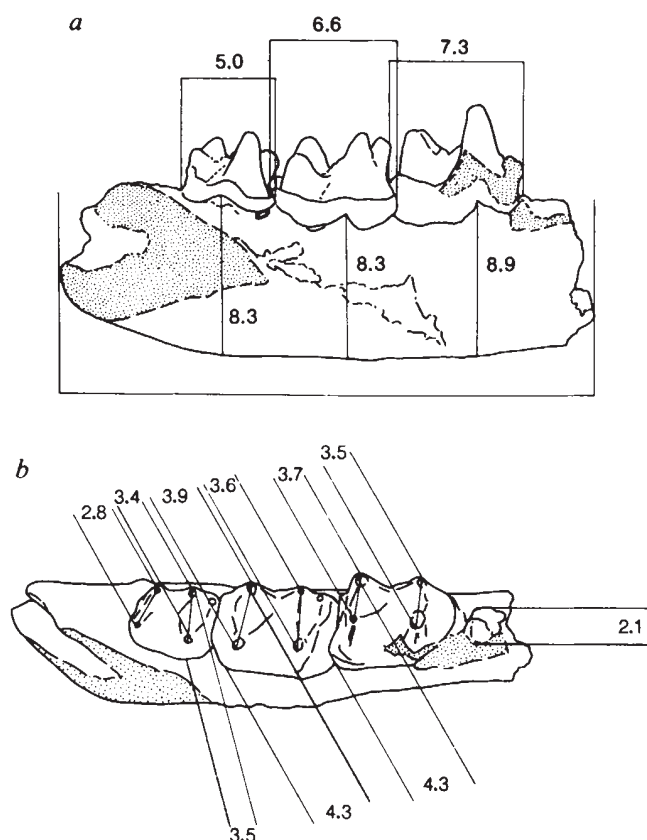
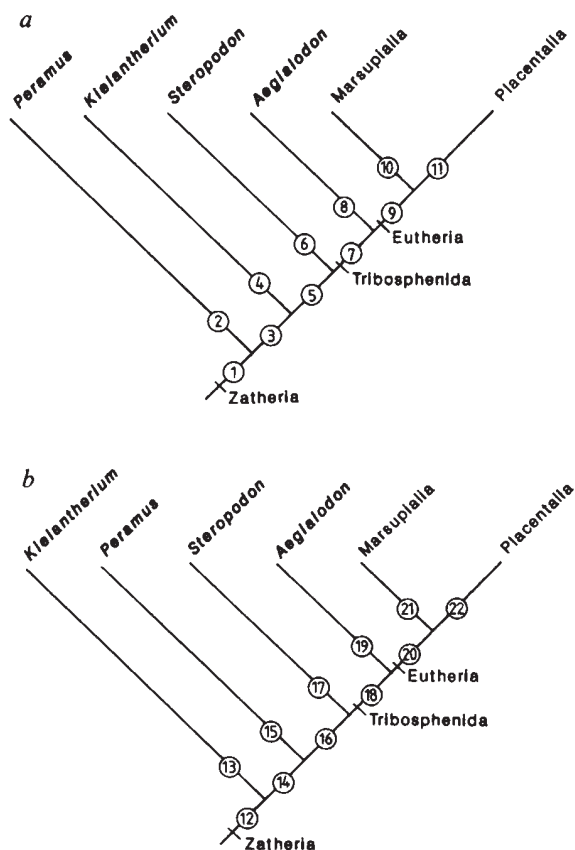


Fig. 2 Measurements (in mm) of the holotype in buccal view (a) and occlusal view (b).



absence of a paraconid on M_1 , high transverse loph-like trigonid and talonid crests, very large talonid, lack of a hypoconulid, and prominent anterior, posterior and buccal cingula are distinctive features that in combination also occur only in the isolated lower molars of the middle Miocene monotreme *Obdurodon insignis*^{3,13}. Further, the dental formula of this Lightning Ridge mammal includes three molars, the number evidently also present in *Ornithorhynchus anatinus*, the only living monotreme to retain even vestigial teeth. However, this number of molars may be a symplesiomorphic feature (see below).

The Lightning Ridge monotreme is significant in that it has relevance for interpretations of relationships among the major groups of mammals. At present, the phylogenetic position of monotremes is in considerable doubt. There are three main hypotheses in vogue: (1) monotremes (possibly with multituberculates, haramiyids and triconodonts) are prototherian mammals and as such the sister-group of the therian mammals; (2) monotremes and multituberculates are monophyletic but this group is one of three trichotomous mammal branches, the other two being the triconodonts and the therian mammals; (3) monotremes (and possibly multituberculates) are therian mammals, as opposed to the prototherian triconodonts (with docodonts) and, within Theria, they form either the sister-group of Zatheria (peramurids plus tribosphenids; refs 4, 14, 15; but see also ref. 16) or the sister-group of the Tribosphenida (Aegialodontidae plus Eutheria). We consider that the Lightning Ridge monotreme provides strongest support for the third hypothesis. Features of *S. galmani* that suggest monophyly with tribosphenid therian mammals include the apparent loss of postdentary bones, the wide and enclosed talonid basins, the enlarged entoconids and the well-developed pre-entocristids (Fig. 3).

Because of uncertainty about the molar number of the peramurid *Peramus tenuirostris*¹⁶⁻¹⁸, we are unable to decide whether peramurids or the newly described *Kielantherium*

Fig. 3 The most-supported cladograms relating *Steropodon galmani*, the oldest and most plesiomorphic monotreme, to other groups of zatherian mammals. The possible relationships of monotremes to multituberculates are noted in the text. Cladogram a presumes that *Peramus tenuirostris* has four adult molars; b presumes that it has three. Potential synapomorphies are as follows: (1) reduction to four adult molars, incipiently basined talonids, stylocones reduced, anterior cingula of lower molars reduced, three talonid cusps; (2) linguall open talonid basins, premolariform first adult molar; (3) enclosed talonid basins with well-developed entocristids; (4) reduction of the entoconids as discrete cusps along the entocristids; (5) loss of postdentary bones from the lower jaw, ? loss of the meckelian groove, wide talonid basins, enlarged entoconids; (6) loss of one molar, paraconid lost from first adult molar, anterior cingula enlarged on lower molars, very wide talonid basins, hypertrophied entoconids and hypoconids, extensive pre-entocristids, loss of hypoconulids, transverse prevallid and postvallid shear; (7) protocone developed; (8) marked reduction of anterior cingula on the lower molars; (9) very wide talonid basins, enlarged protocones, transverse prevallid and postvallid shear, reduction of the postmetacristids; (10) loss of deciduous premolars, enlarged stylar cusp C, enlarged metacones, approximated entoconids and hypoconulids; (11) loss of one adult molar, reduction of stylar cusps, molariform dP4; (12) reduction to four molars, incipiently basined talonids, two talonid cusps (the hypoconid and hypoconulid), stylocones reduced, anterior cingula of lower molars reduced; (13) basined talonids with extensive pre-entocristids; (14) loss of one adult molar, entoconids developed, ? loss of meckelian groove; (15) submolariform posterior premolar, small lingual talonid basins; (16) loss of postdentary bones from the lower jaw, completely enclosed and wide talonid basins, well-developed pre-entocristids; (17) loss of the paraconid from the anterior adult molar, enlarged anterior cingula on the lower molars, hypertrophied entoconids and hypoconulids, loss of hypoconulids, transverse prevallid and postvallid shear; (18) protocone developed, pre-entocristids reduced; (19) marked reduction of anterior cingula on the lower molars; (20) same as (9); (21) same as (10) plus supernumerary molars (normally there are four adult molars in marsupials as well as a deciduous first molar); and (22) reduction of stylar cusps, molariform dP4.

*gobiensis*¹⁶ represents the plesiomorphic sister-group of monotremes plus tribosphenids. While Dashzeveg and Kielan-Jaworowska¹⁶ suggest that *K. gobiensis* is a tribosphenid similar to *Aegialodon dawsoni*, the presence in the latter but not the former of three well-developed talonid cusps and a wide talonid basin seems to us to be sufficient reason to question the inclusion of both forms within the same family. Pending the results of studies of monotreme occlusion, we suggest that the structure of monotremes (as evidenced by the albeit autapomorphically specialized *S. galmani*) represents an evolutionary stage between these two zatherians (Fig. 3).

The fact that distinctive monotremes were present in Australia in the early Cretaceous indicates not only the antiquity of this group but also that by this time at least some elements of Australia's fauna were exhibiting regional endemism of the sort that characterizes much of the continent's known Cenozoic faunas⁹. However, the paucity of knowledge about the early Cretaceous mammals of other continents means that we cannot exclude the possibility that monotremes also occurred on other continents in the Cretaceous.

One additional curious aspect of *S. galmani* is its large size. Not only is it the largest mammal known from the early Cretaceous, it may be one of the two largest Mesozoic mammals known, the other being the badger-sized late Cretaceous stegodontid *Didelphodon vorax*¹⁹. Overlap in body size between Mesozoic mammals and dinosaurs has not previously been recorded²⁰. The Lightning Ridge local fauna, however, includes some very small hysilophodontid dinosaurs which were very little if at all larger than *S. galmani*.

Finally, the Lightning Ridge monotreme is heraldic because it comes from a continent that so far has failed to produce any pre-Miocene terrestrial mammals². As such, it represents a long-overdue reassurance that Australia did have an early Tertiary mammal record and that perseverance in the search will be worthwhile.

We thank Esso Australia, the State Government of New South Wales and the Director and Trustees of the Australian Museum for raising or providing funds necessary to acquire the Galman collection for the Australian Museum. We also acknowledge the contribution made by David and Alan Galman who bought and assembled a unique collection of opalized fossils from Lightning Ridge, including the *S. galmani* dentary. Michael Plane and Brian Senior helped in interpreting the geological setting of the specimen. Robert Jones helped with preparation of the specimen and John Fields, Alex Ritchie and Jenny Taylor produced the photographs. Colleagues who also contributed thoughts about the significance of the specimen include Richard Tedford, Michael Plane, Michael Woodburne, Neville Pledge, Thomas Rich, Suzanne Hand, Lyndal Dawson and Kenneth Aplin. Zophia Kielan-Jaworowska provided a cast of the important *K. gobiensis*.

Received 15 April; accepted 1 October 1985.

1. Tedford, R. H., Banks, M. R., Kemp, N. R., McDougall, I. & Sutherland, F. L. *Nature* **225**, 141-142 (1975).
2. Archer, M. & Hand, S. in *Vertebrate Zoogeography and Evolution in Australasia* (eds Archer, M. & Clayton, G.) 517-565 (Hesperian, Perth, 1984).
3. Woodburne, M. O. & Tedford, R. H. *Novit. zool.* **2588**, 1-11 (1975).
4. Kemp, T. S. *Zool. J. Linn. Soc.* **77**, 353-384 (1983).
5. Hawke, J. M. & Bourke, D. J. *Bull. geol. Surv. N.S.W.* **31**, 23-123 (1984).
6. Molnar, R. E. *Mém. Soc. géol. Fr.* **139**, 131-143 (1980).
7. Gleadow, A. J. W. & Duddy, I. R. in *Gondwana Five* (eds Cresswell, M. M. & Vella, P.) (Belkema, Rotterdam, 1980).
8. Byrnes, J. G. *Geol. Surv. N.S.W. Rep. GS1977/005*, 1-17 (1977).
9. Morgan, R. *Bull. geol. Surv. N.S.W.* **31**, 143-160 (1984).
10. Molnar, R. E. in *The Fossil Vertebrate Record of Australasia* (eds Rich, P. V. & Thompson, E. M.) 170-225 (Monash University, Melbourne, 1982).
11. Molnar, R. E. in *Vertebrate Zoogeography and Evolution in Australasia* (eds Archer, M. & Clayton, G.) 337-341 (Hesperian, Perth, 1984).
12. Archer, M. in *Vertebrate Zoogeography and Evolution in Australasia* (eds Archer, M. & Clayton, G.) 585-625 (Hesperian, Perth, 1984).
13. Archer, M., Plane, M. D. & Pledge, N. S. *Aust. Zool.* **20**, 9-27 (1979).
14. Crompton, A. W. & Jenkins, F. A. Jr in *Mesozoic Mammals the First Two-Thirds of Mammalian History* (eds Lillegraven, J. A., Kielan-Jaworowska, Z. & Clemens, W. A.) 7-58 (University of California Press, 1979).
15. Prothero, D. R. *Bull. Am. Mus. nat. Hist.* **167**, 277-362 (1981).
16. Dashzeveg, D. & Kielan-Jaworowska, Z. *Zool. J. Linn. Soc.* **8**, 217-227 (1984).
17. McKenna, M. C. in *Phylogeny of the Primates* (eds Luckett, W. P. & Szalay, F. S.) 21-46 (Plenum, New York, 1975).

18. Clemens, W. A. & Mills, J. R. E. *Bull. Br. Mus. nat. Hist.* **20**, 89-113 (1971).
19. Clemens, W. A. in *Mesozoic Mammals the First Two-Thirds of Mammalian History* (eds Lillegraven, J. A., Kielan-Jaworowska, Z. & Clemens, W. A.) 192-220 (University of California Press, 1979).
20. Lillegraven, J. A. in *Mesozoic Mammals the First Two-Thirds of Mammalian History* (eds Lillegraven, J. A., Kielan-Jaworowska, Z. & Clemens, J. A.) 259-276 (University of California Press, 1979).

Kinship, reciprocity and synergism in the evolution of social behaviour

David C. Queller

Department of Biology, Rice University, PO Box 1892, Houston, Texas 77251, USA

There are two ways to model the genetic evolution of social behaviour. Population genetic models using personal fitness¹⁻⁹ may be exact and of wide applicability, but they are often complex and assume very different forms for different kinds of social behaviour. The alternative, inclusive fitness models¹⁰⁻¹², achieves simplicity and clarity by attributing all fitness effects of a behaviour to an expanded fitness of the actor. For example, Hamilton's rule states that an altruistic behaviour will be favoured when $-c + rb > 0$, where c is the fitness cost to the altruist, b is the benefit to its partner, and r is their relatedness. But inclusive fitness results are often inexact for interactions between kin¹⁻⁵, and they do not address phenomena such as reciprocity¹³⁻¹⁵ and synergistic effects^{7,8,16} that may either be confounded with kinship or operate in its absence. Here I develop a model the results of which may be expressed in terms of either personal or inclusive fitness, and which combines the advantages of both; it is general, exact, simple and empirically useful. Hamilton's rule is shown to hold for reciprocity as well as kin selection. It fails because of synergistic effects, but this failure can be corrected through the use of coefficients of synergism, which are analogous to the coefficient of relatedness.

For simplicity, assume that all members of a diploid population interact in pairs, that each may or may not have an opportunity to aid its partner and that the genetic component of altruism is due to two alleles at a single locus, with no overdominance. Table 1 defines the random variables used in the model. Each is defined over the whole population. The phenotypic

Table 1 Random variables used in the model

- G = an individual's frequency of the altruism allele (0, $\frac{1}{2}$ or 1)
 P = an individual's phenotypic value (1 when altruistic, 0 when not)
 W = total fitness
 W_0 = fitness if aid is neither given nor received
 C = fitness cost if individual were to be altruistic (that is, if $P = 1$)
 B = fitness benefit if partner were to be altruistic (that is, if $P' = 1$)
 D = deviation from additivity of C and B if both were altruistic (that is, if $PP' = 1$)
 G', P', B', D' = an individual's partner's values of G, P, B, D

values (P, P') are defined as either 1 or 0, where, in contrast to inclusive fitness models, the latter case applies to all individuals who did not perform the altruism, including those who had no opportunity to do so. Fitness components are also defined for all individuals, for example, C is defined, even for a non-altruist, as the cost it would incur if it were altruistic. Finally, assume that each fitness component is distributed independently of G, G', P and P' .

An individual's fitness, W , can be written as

$$W = W_0 - CP + BP' + DPP' \quad (1)$$

where multiplication by the phenotypic values ensures that the fitness components are counted only when the appropriate behaviours are performed. Assuming a fair meiosis, Price's selection equation¹⁷ shows that $\Delta G = \text{Cov}(G, W)/\bar{W}$, so the

altruism allele is favoured ($\Delta \bar{G}$ positive) when $\text{Cov}(G, W) > 0$. Substituting for W from equation (1) and splitting into four terms yields

$$\text{Cov}(G, W_0) - \text{Cov}(G, CP) + \text{Cov}(G, BP') + \text{Cov}(G, DPP') > 0 \quad (2)$$

Since W_0 , C , B and D are distributed independently of G , G' , P and P' , the first term is zero and the remainder may be rewritten as $-\bar{c}\text{Cov}(G, P) + \bar{b}\text{Cov}(G, P') + \bar{d}\text{Cov}(G, PP') > 0$, where $\bar{c}$, $\bar{b}$ and $\bar{d}$ are the sample means of C , B and D .

The mean fitness effects have now been extracted from the covariances that embody the population structure. Dividing through by $\text{Cov}(G, P)$ yields:

$$-\bar{c} + \bar{b} \cdot \frac{\text{Cov}(G, P')}{\text{Cov}(G, P)} + \bar{d} \cdot \frac{\text{Cov}(G, PP')}{\text{Cov}(G, P)} > 0 \quad (3)$$

This is still a personal fitness formulation since all fitness effects are assigned to the individual affected rather than to the one responsible for the effect. An inclusive fitness form is obtained by rewriting condition (2) as $\text{Cov}(G, W_0) + \text{Cov}(G, CP) + \text{Cov}(G', B'P) + \text{Cov}(G', D'P'P) > 0$. The same set of interactions are scored for genotype, phenotype and fitness effect. In the third and fourth terms we have simply switched the arbitrary labels designating when an individual is considered as self and when as someone else's partner. Mathematically this amounts to reordering the terms within the covariance, leaving the total unchanged. Proceeding as before now yields

$$-\bar{c}' + \bar{b}' \cdot \frac{\text{Cov}(G', P)}{\text{Cov}(G, P)} + \bar{d}' \cdot \frac{\text{Cov}(G', P'P)}{\text{Cov}(G, P)} > 0 \quad (4)$$

We now have equivalent results from the viewpoints of personal fitness (condition 3) and inclusive fitness (condition 4). When costs and benefits are additive, $\bar{d} = \bar{d}' = 0$, and the third terms disappear. What remains is Hamilton's rule, using covariance measures of relatedness previously proposed by Orlove and Wood¹⁸ (condition 3) and by Michod and Hamilton¹⁹ (condition 4). Here, however, these measures are taken over all individuals in the population, not just those with an opportunity to perform altruism, and this minor change makes the two measures equivalent.

These relatedness coefficients actually measure more than pedigree connections. The derivation is unchanged for altruism expressed in a conditional manner: be altruistic to an untested partner; otherwise, behave towards the partner as he behaved towards you. This reciprocal altruism strategy can be selected for the same reason as altruism towards kin: covariance between the performance of the behaviour and the recipient's frequency of the altruism allele. With reciprocal altruism a positive covariance arises when a partner's prior behaviour serves as a reliable indicator of possession of the reciprocity genes.

The third terms of conditions (3) and (4) correct Hamilton's rule for non-additivity. They consist of the product of the average deviation effect, $\bar{d}$, and a population structure coefficient. I will call these s coefficients, or coefficients of synergism, because they depend on the joint behaviour of both partners (PP'). In every other respect they are the same as the r coefficients measuring relatedness and reciprocity; both depend on a covariance between the behaviour(s) causing a fitness effect and the genes of the individual receiving it. These s coefficients are positive, even when interactants are not relatives (though they approach zero for very low P and P'), so that the sign of $\bar{d}$ alone determines the direction of error in Hamilton's rule. Since relatedness is not required, the third term is best viewed as being due to something other than kin selection: a form of trait group selection⁹ or, to use a more evocative term, synergistic selection¹⁶. In this view, Hamilton's rule is correct for pure kin selection (and, as shown above, reciprocal altruism); its failure is caused by a distinct process.

A variety of biological phenomena can be represented by the D term²⁰. Several examples are given below. In each, D represents some combined effect of two altruists, acting either

sequentially or simultaneously, beyond their summed effects if they acted alone.

For behaviours in which costs and benefits take the form of successively altered survival probabilities, such as warning calls, a multiplicative fitness model is appropriate²¹. If we let the base fitness be unity, the fitness of an individual receiving both cost and benefit is not $1 - C + B$, but $(1 - C)(1 + B)$ or $1 - C + B - BC$, so that $D = -BC$. The negative deviation from additivity means that altruism will be harder to select than predicted by Hamilton's rule, as other models have shown^{4,5}. In effect, benefits are squandered when the recipient dies because of its own altruism.

A distasteful insect with warning coloration runs an increased risk of being seen and eaten by a predator. This cost may be compensated if the bad taste causes the predator to avoid its victim's neighbours²². A general avoidance of neighbours, regardless of their coloration, would be represented by B , and this could select for warning coloration if neighbours tend to be related. But any additional avoidance of only those neighbours with warning coloration must be represented by a positive D because this gain applies only to those who play the same strategy as the altruist.

Male lions may cooperate in taking over or retaining a pride²³. In searching for a partner, a cooperative lion may incur some cost, C , whether or not the potential partner actually cooperates. The fitness gain, however, applies only if cooperation does ensue ($PP' = 1$), and is therefore represented not by B , but by a synergistic D . For both lion cooperation and warning coloration, the positive D means Hamilton's rule underestimates the ease with which these traits are selected. In fact, if the altruistic trait can reach non-zero levels by drift, the s coefficient will be positive even in the absence of relatedness, so that a positive D could sometimes select for altruism by itself.

The model is quite general. It is not restricted to altruistic behaviours, since B and C may be assigned negative values. Additional synergistic effects can be incorporated by adding terms to the fitness decomposition in equation (1). For example, if three altruists ($P = P' = P'' = 1$) cause some additional effect, E , beyond that caused by two, the added term would be $EPP'P''$. The starting point of the model, Price's equation, holds for any mating system and is easily adjusted for haplodiploidy. It is also valid for any linear function of any number of alleles at any number of loci¹⁷. Thus, G and G' may be interpreted as the breeding value for the altruistic behaviour. Since this value can be estimated through breeding experiments²⁴, the model should be useful in empirical studies.

This work was supported in part by a NATO postdoctoral fellowship at the University of Sussex. For discussion and advice I thank B. Charlesworth, W. D. Hamilton, C. R. Hughes, J. Maynard Smith, R. E. Michod, J. Seger, J. Strassmann, M. Taper and M. J. Wade. Special thanks to M. F. Willson for logistical support and P. D. Taylor for help in simplifying the derivation.

Received 22 April; accepted 24 September 1985.

- Boorman, S. A. & Levitt, P. R. *The Genetics of Altruism* (Academic, New York, 1980).
- Charlesworth, B. *J. theor. Biol.* **72**, 297-319 (1978).
- Templeton, A. R. *Am. Nat.* **114**, 515-524 (1979).
- Cavalli-Sforza, L. L. & Feldman, M. W. *Theor. popul. Biol.* **14**, 268-280 (1978).
- Uyenoyama, M. K. & Feldman, M. W. *Am. Nat.* **120**, 614-627 (1982).
- Wade, M. J. *Am. Nat.* **125**, 61-73 (1985).
- Matessi, C. & Jayakar, S. D. *Theor. popul. Biol.* **9**, 360-387 (1976).
- Cohen, D. & Eschel, I. *Theor. popul. Biol.* **10**, 276-302 (1976).
- Wilson, D. S. *The Natural Selection of Populations and Communities* (Benjamin/Cummings, Menlo Park, 1980).
- Hamilton, W. D. *J. theor. Biol.* **7**, 1-52 (1964).
- Seger, J. *J. theor. Biol.* **91**, 191-213 (1981).
- Michod, R. E. *A. Rev. ecol. Syst.* **13**, 23-55 (1982).
- Trivers, R. L. *Q. Rev. Biol.* **46**, 35-57 (1971).
- Axelrod, R. & Hamilton, W. D. *Science* **211**, 1390-1396 (1981).
- Brown, J. S., Sanderson, M. J. & Michod, R. E. *J. theor. Biol.* **99**, 319-339 (1982).
- Maynard Smith, J. in *Current Problems in Sociobiology* (eds. King's College Sociobiology Group) 29-44 (Cambridge University Press, 1982).
- Price, G. R. *Nature* **227**, 520-521 (1970).
- Orlove, M. J. & Wood, C. L. *J. theor. Biol.* **73**, 679-686 (1978).
- Michod, R. E. & Hamilton, W. D. *Nature* **288**, 694-697 (1980).
- Queller, D. C. *Biol. J. Linn. Soc.* **23**, 133-143 (1984).
- Maynard Smith, J. *Theor. popul. Biol.* **18**, 151-159 (1980).
- Harvey, P. H., Bull, J. J., Pemberton, M. & Paxton, R. J. *Am. Nat.* **119**, 710-719 (1982).
- Bygott, J. D., Bertram, B. C. R. & Hanby, J. P. *Nature* **282**, 839-841 (1979).
- Falconer, D. S., *Introduction to Quantitative Genetics* 2nd ed (Longman, London 1981).

Plasticity of functional epithelial polarity

George J. Schwartz*, Jonathan Barasch†
& Qais Al-Awqati†

* Departments of Pediatrics and Physiology/Biophysics, Albert Einstein College of Medicine, Bronx, New York 10461, USA, and † Departments of Medicine and Physiology, Columbia University, College of Physicians and Surgeons, New York, New York 10032, USA

The fundamental characteristics that allow vectorial transport across an epithelial cell are the differential sorting and insertion of transport proteins either in the apical or the basolateral plasma membrane, and the preferential association of endocytosis and exocytosis with one or the other pole of the cell. Asymmetrical cellular structure and function, being manifestations of terminal differentiation, might be expected to be predetermined and invariant. Here we show that the polarity of transepithelial H^+ transport, endocytosis and exocytosis in kidney can be reversed by environmental stimuli. The HCO_3^- -secreting cell in the cortical collecting tubule is found to be an intercalated cell possessing a Cl^-/HCO_3^- exchanger in the apical membrane and proton pumps in endocytic vesicles that fuse with the basolateral membrane; the H^+ -secreting cell in the medullary collecting tubule has these transport functions on the opposite membranes. Further, the HCO_3^- -secreting cell can be induced to change its functional polarity to that of the H^+ -secreting cell by acid-loading the animal.

We identified a subset of cells in isolated cortical collecting tubules using the mitochondrial accumulation of the fluorescent cation 3,3'-dipentylloxycarbocyanine (Di-O-C₅(3)) (Fig. 1a). These mitochondrion-rich cells are enriched in carbonic anhydrase and transport protons^{1,2}; that is, they are the intercalated cells. We stained cortical collecting tubules with 6-carboxyfluorescein diacetate; this is a permeant nonfluorescent precursor of the pH-sensitive dye 6-carboxyfluorescein which is released by cytoplasmic esterases such as carbonic anhydrase, so these cells will preferentially accumulate the dye (Fig. 1b). There were 100–150 intercalated cells per mm of tubule, accounting for about 25% of all the cells in this segment (L. M. Satlin and G.J.S., unpublished observations). The cytoplasmic pH (pH_i) of the intercalated cell, measured by excitation ratio microspectrofluorometry²⁻⁴, was on average 0.4 pH units higher than that of the principal cell. When collecting tubules were perfused with fluorescent macromolecules, only 5–10% of the intercalated cells in the cortical tubule, but most of those in the medullary segment, endocytosed the fluorophores (Fig. 1c–e)³. Since the medullary collecting tubule secretes H^+ whereas the cortical collecting tubule secretes HCO_3^- (Table 1), it is likely that the intercalated cell showing luminal endocytosis is the H^+ -secreting cell, while that lacking it secretes HCO_3^- .

To test this directly, we removed chloride from the basolateral solution and measured pH_i in individual intercalated cells simultaneously identified as those showing or not showing endocytosis by the uptake of rhodamine-bovine serum albumin (BSA) (see Table 1 legend). Removal of basolateral chloride stimulated HCO_3^- secretion in the cortical collecting tubule and inhibited H^+ secretion (HCO_3^- absorption) in the medullary collecting tubule (Table 1). It also increased the cell pH_i by 0.2 ± 0.1 units in both cortical and medullary intercalated cells that had endocytosed rhodamine-BSA from the lumen (Table 1). The increased cellular pH_i and decreased H^+ secretion in medullary collecting tubules indicate that the luminally endocytosing cell has a basolateral Cl^-/HCO_3^- exchanger and is the H^+ -secreting cell. In the cells that did not endocytose rhodamine-BSA, removal of basolateral Cl^- decreased cell pH_i by 0.4 ± 0.1 pH units (Table 1). As this manoeuvre stimulated HCO_3^- secretion, the reduction in pH_i indicates that this cell is the HCO_3^- -secreting cell, with a Cl^-/HCO_3^- exchanger located in the luminal membrane.

Table 1 Net HCO_3^- transport ($J_{HCO_3^-}$) and cell pH_i (pH_i) in individually identified intercalated cells of isolated perfused cortical and medullary collecting tubules

	$J_{HCO_3^-}$ ($\mu\text{mol min}^{-1}$ mm^{-1})	Luminal endocytosis	pH_i No luminal endocytosis
Cortical collecting tubule			
Control	-3.3 ± 1.6 $n = 13$	7.23 ± 0.06 $n = 7$	7.34 ± 0.09 $n = 7$
Cl^- -free bath	$-12.0 \pm 2.3^{\dagger}$	$7.40 \pm 0.09^*$	$6.96 \pm 0.13^*$
Medullary collecting tubule			
Control	9.1 ± 1.6 $n = 5$	7.14 ± 0.07 $n = 5$	—
Cl^- -free bath	$2.1 \pm 1.6^*$	$7.33 \pm 0.04^{\dagger}$	—

One to four cells were studied in each perfused tubule to provide a mean for each animal. Net fluid absorption in four cortical collecting tubules perfused and bathed in identical solutions was negligible (control = 0.023 ± 0.024 ; Cl^- -free bath = 0.022 ± 0.027 $\text{nl min}^{-1} \text{mm}^{-1}$, $P > 0.5$). The tubules were perfused at $1-2$ $\text{nl min}^{-1} \text{mm}^{-1}$. Bicarbonate concentration in nl samples was measured by microcalorimetry¹⁰ and net bicarbonate transport was calculated from the product of the difference in HCO_3^- concentration between perfused and collected fluids and the flow rate, and normalized to mm of tubular length. All experiments were done at 38°C and used an artificial solution (290 ± 1 mosmol per kg) containing (in mM): NaCl (114), NaHCO_3 (25), K_2HPO_4 (2.5), CaCl_2 (2.0), MgSO_4 (1.2), glucose (5.5), L-alanine (6.0), Na-lactate (4.0) and citrate (1.0); it was gassed with 95% $\text{O}_2/5\%$ CO_2 to yield a pH of 7.4. At least three collections of tubular fluid were obtained for measurement of total CO_2 . Plus values denote bicarbonate absorption (H^+ secretion) while negative values denote bicarbonate secretion. Cell pH_i was measured in individually identified cells using 20–30 μM 6-carboxyfluorescein diacetate in the bath to load the cells for 5 min with 6-carboxyfluorescein. Simultaneously, the lumen was perfused with the same artificial solution containing rhodamine-BSA as a fluorescent macromolecule at $0.5-1$ mg ml^{-1} . The cells that endocytosed the fluorescent macromolecule, considered to be H^+ -secreting cells, appeared red under the rhodamine filter cassette (Fig. 1d, e) and bright green under the fluorescein filter block (Fig. 1b). Those cells which did not endocytose rhodamine from the lumen appeared bright green under the fluorescein filter cassette and showed no red staining with the rhodamine filters. Each cell was identified as rhodamine-positive or -negative and centred in the microscope photometer aperture. The filter cassette was then changed to fluorescein, the single cell was excited at 450 and 490 nm, and the emission was recorded at 535 nm in a Farrand microspectrofluorometer via the photometer attached to the camera port of the inverted microscope. As there were so few cells that did not endocytose rhodamine-BSA in the medullary collecting duct, we did not study such cells in that segment. Because the pH sensitivity of 6-carboxyfluorescein tends to saturate at pH values > 7.3 , these isolated tubules were studied at a $p\text{CO}_2$ of 50–55 mm Hg. This mild respiratory acidosis slightly decreased the cell pH_i . At the end of each experiment, the tubule was exposed to 100 μM of the electrogenic proton ionophore CCCP (carbonyl cyanide *p*-chlorophenyl hydrazone) in 150 mM KCl plus 10 mM phosphate buffer. This combination depolarized the cell and clamped the pH of the cytoplasm to that of the bath. The pH of the bath was compared with that measured fluorimetrically in 5–15 cells of the clamped tubules. The recorded pH values were obtained from a calibration curve over a pH range of 5.5–7.5 in the same calibration buffer. The fluorescence intensity of the 6-carboxyfluorescein-containing extracellular calibration solution was chosen to be approximately the same as the intensity of the cells studied *in vitro*, $\sim 1-2$ μM . The average pH difference between the fluorimetric cellular pH_i and the pH of the bath as measured by a microelectrode was 0.02 ± 0.03 units ($n = 10$ tubules) and was not a significant difference.

* $P < 0.05$; † $P < 0.01$ (Cl^- -free bath versus control). Mean $\pm$ s.e., n = number of animals.

Secretion of HCO_3^- by a luminally located Cl^-/HCO_3^- exchanger implies that a proton-translocating H^+ -ATPase exists at the basolateral membrane. Such proton pumps are packaged in endocytic vesicles, as we have shown in the H^+ -secreting cells²⁻⁴. To demonstrate basolateral endocytosis into acid vesicles, fluoresceinated dextran (relative molecular mass 70,000) was added to the basolateral medium of collagenase-treated cortical collecting tubules. Some, but not all, intercalated cells endocytosed the macromolecules into vesicles (Fig. 1f) that maintained an intravesicular pH of 6.0 ± 0.2 ($n = 14$ tubules, with 5–6 cells measured per tubule), similar to that found previously in cells endocytosing fluorescein-dextran from the lumen^{2,3}. By measuring total cellular fluorescence in 6 tubules (4–5 cells studied per tubule), we found that increasing ambient $p\text{CO}_2$ induced the exocytotic release of $40 \pm 3\%$ of the basolaterally endocytosed fluorescein-dextran (or the pH -insensitive dye Lucifer yellow), similar to results found for the luminally endocytosing cell³.

To confirm the presence of basolateral endocytosis, we exposed collagenase-treated cortical collecting tubules to horse-

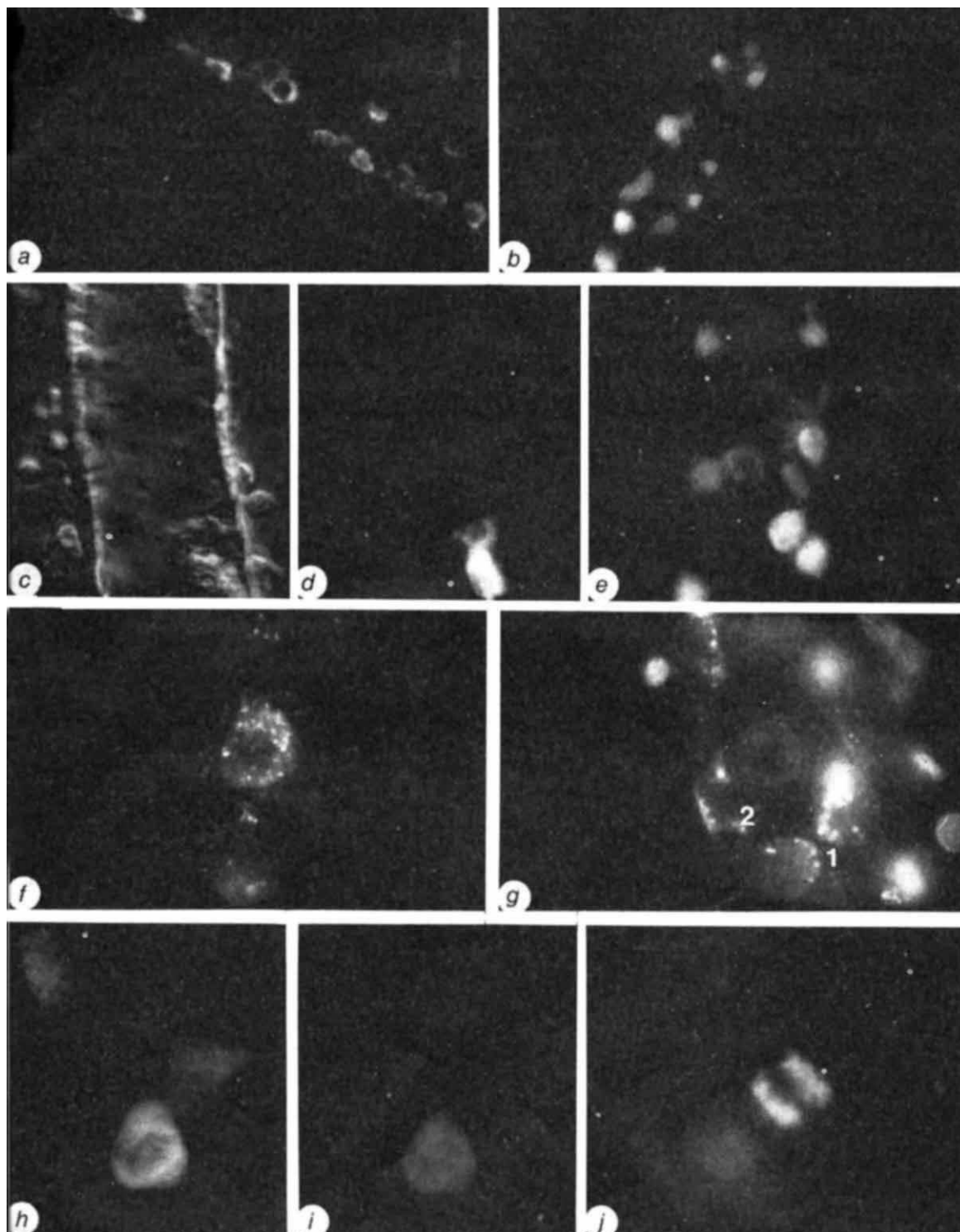


Fig. 1 *a*, Isolated cortical collecting tubule incubated for 5 min in 50 nM Di-O-C₅(3), a fluorescent cation which accumulates in mitochondria. At this low concentration of the dye the reticular, perinuclear pattern is seen only in some cells, the mitochondrion-rich (intercalated) cells $\times 200$. *b*, Isolated cortical collecting tubule incubated for 5 min in 20 μ M 6-carboxy fluorescein diacetate. Bright fluorescence is seen in the mitochondrion-rich cells. These cells are enriched in carbonic anhydrase, a potent esterase, and thus preferentially take up this dye. Further, as these cells are more alkaline, the dye will appear brighter than in the other cells. $\times 200$. *c, d*, Cortical collecting tubules perfused for 10 min with rhodamine-BSA (1 mg ml⁻¹) and then washed free of luminal fluorophore. In *c*, using Hoffman modulation contrast optics, note the presence of intercalated cells with a characteristic bulging surface at the top right and bottom right. *d*, The same field using epifluorescence optics: only the bottom right cells have internalized the fluorophore. The pattern of fluorescence is diffuse, with negative nuclear staining. Because of the enrichment of the cells with vesicles and consequent light scattering, it is difficult to resolve individual vesicles. $\times 300$. *e*, Fluorescence micrograph of a medullary collecting tubule perfused in a similar manner to that of *d*. Endocytosis of the marker is seen in many more cells than in the cortical collecting tubule in *d*. $\times 200$. *f*, Isolated perfused cortical collecting tubule incubated for 20 min with 3 mg ml⁻¹ Lucifer yellow in the basolateral bathing medium. After extensive washing the tubule was mounted between two coverslips and photographed. A punctate uptake of this small non-permanent dye is seen in some cells. $\times 500$. *g*, Isolated cortical collecting tubule incubated for 2 min in 10 μ M acridine orange. This weak base is concentrated in acid vesicles and its emission spectrum undergoes a redshift at high concentration. The cell numbered 1 shows a concentration of sub-apical vesicles with no basolateral acid vesicles, suggesting that it is the H⁺-secreting cell. Cell 2 (probably a HCO₃⁻-secreting cell) has many sub-apical and basolateral vesicles. Addition of 5 μ M nigericin, a K⁺H⁺ exchange ionophore, abolished the orange staining in the vesicles, indicating that the dye accumulated in these vesicles because of the presence of a pH gradient across their membranes. $\times 500$. *h, i*, A cortical collecting tubule perfused at 38 °C for 10 min with rhodamine-BSA (1 mg ml⁻¹), washed free of the dye, cooled to 9 °C for 3 min and perfused with fluorescein-conjugated peanut lectin (0.5 mg ml⁻¹). After washout of fluorescein lectin, the tubule was mounted between two coverslips and photographed at $\times 500$. *h*, Note two patterns of fluorescence, one appearing diffuse with a negative nuclear stain while the other shows sharper edges characteristic of surface binding. *i*, The same field observed with rhodamine filters. Note endocytosis of the macromolecules only in the cell that shows a diffuse pattern and negative nuclear stain in *h, j*, Collecting tubule isolated from the ampulla of the newborn rabbit kidney, fixed with 2% glutaraldehyde in 0.1 M cacodylate buffer (pH 7.4) containing 10 μ g ml⁻¹ Hoechst 33258, the fluorescent chromatin-binding stain. After 1 h of exposure the segment was mounted between two coverslips. Note that the chromosomes are condensed in the polar positions of anaphase. $\times 500$. Tubules isolated as described in Table 1.

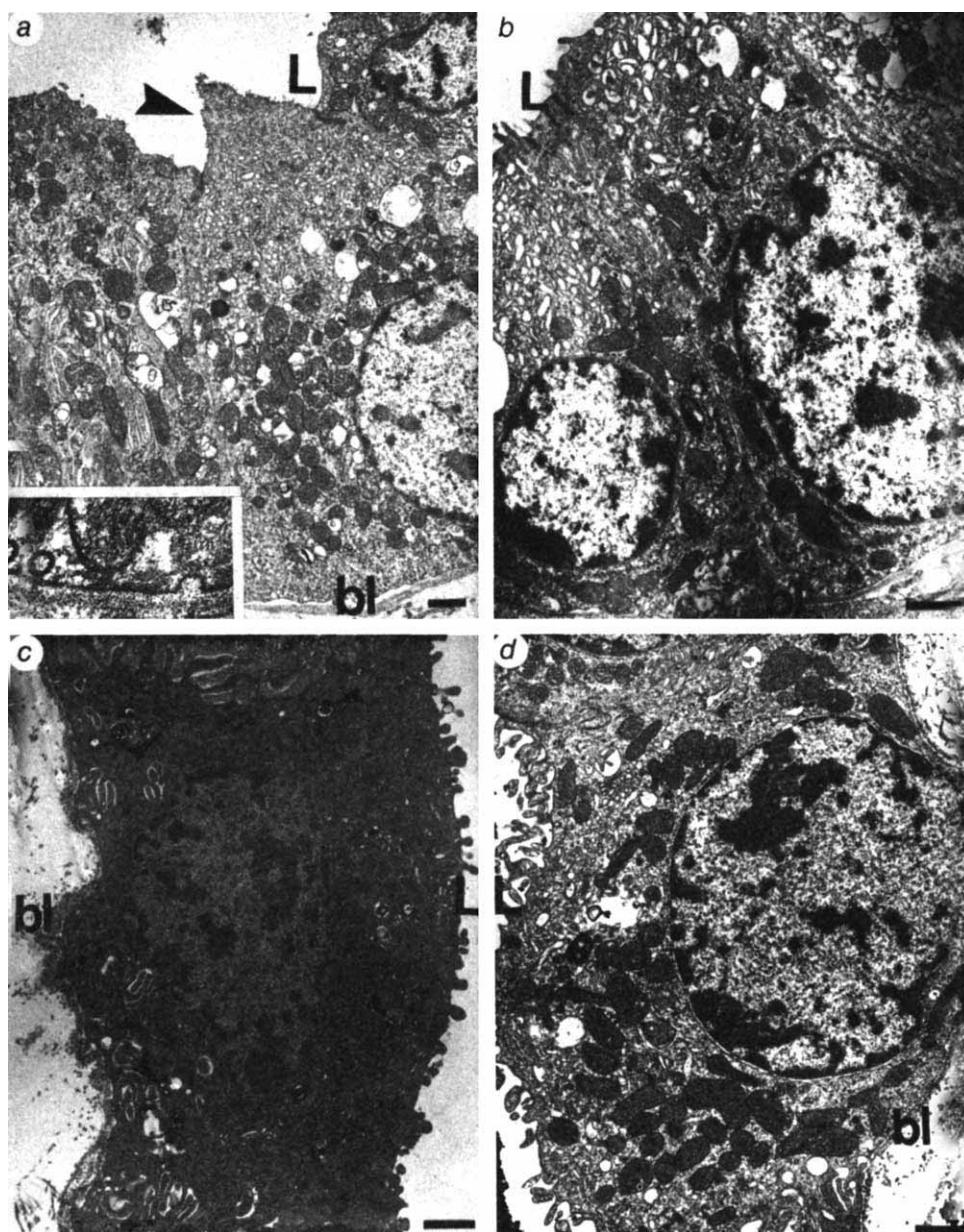


Fig. 2 Electron microscopy of cortical collecting tubules in intact kidneys (*a, b*) and in isolated tubules incubated for 10 min in 2% horseradish peroxidase (type VI, Sigma) (*c, d*). Tubules were fixed with 4% glutaraldehyde and post-fixed with 1% OsO_4 . They were then reacted with the diaminobenzidine/glucose/glucose oxidase system of Itoh *et al.*¹¹ for 30 min after pretreatment with 0.1% CoCl_2 . Two types of intercalated (mitochondria-rich) cells are seen. In *a* and *c* apical vesicles are surmounted by a dense cytoskeletal meshwork (*a*, arrowhead) and coated endocytic vesicles and omega figures are found along the basolateral membrane (*a*, inset). In *b* and *d* the apical cytoplasm is less dense and fusions between the apical vesicles and luminal membrane can be seen. Perfusion of the basolateral surface with horseradish peroxidase in isolated tubules for 10 min (*c, d*) showed considerable basolateral endocytosis in the mitochondria-rich cells having apical cytoskeletal caps (*c*, arrowheads) but no such uptake was found in the other apical vesicle-rich cell (*d*). Longer periods of exposure to horseradish peroxidase resulted in uptake of the marker in all cells. Hence, cell *c* has a higher rate of basolateral endocytosis than the other cells. bl, Basolateral; L, luminal. Scale bar, 1 μm ; inset in *a*, $\times 48,000$.

radish peroxidase from the basolateral side only. Electron microscopy revealed at least two types of intercalated cells⁵ (Fig. 2). Only one type showed basolateral uptake of horseradish peroxidase (compare Fig. 2*c* and *d*). The cells that endocytosed horseradish peroxidase contained large numbers of sub-apical vesicles (Fig. 2*a, c*), separated from the plasmalemma by a dense 'mat' of fibrillar cytoskeletal matrix (Fig. 2*a*). Also apparent were many endo-exocytic (omega) figures and coated vesicles at the basolateral membrane (Fig. 2*a*, inset). In the other type of intercalated cell these omega figures and coated vesicles occurred predominantly at the apical membrane (Fig. 2*b, d*). Such coated vesicles (probably clathrin-coated vesicles) are known to contain the H^+ -ATPase⁶⁻⁸. Both sub-apical and basolateral vesicles were acidic, as shown by the accumulation of the weak base acridine orange (Fig. 1*g*). Cell 2 in Fig. 1*g* is probably the HCO_3^- -secreting cell containing apical and basolateral acid vesicles; cell 1 is probably an H^+ -secreting cell containing a large sub-apical concentration of vesicles. The principal cells did not contain many acid vesicles.

When animals were given a chronic acid load, the transport properties of cortical collecting tubules removed 24 h later and perfused *in vitro* at pH 7.4 were converted from HCO_3^- secretion

to H^+ secretion (Table 2). This was accompanied by a large increase in the number of H^+ -secreting cells (that is, cells showing luminal endocytosis) (Table 2). To test whether the increased H^+ secretion was due to a conversion of the polarity of the HCO_3^- -secreting cell or to a proliferation in the H^+ -secreting cell, we counted both H^+ - and HCO_3^- -secreting cells. Intercalated cells were identified by their ability to bind peanut lectin⁹. When fluorescent peanut lectin was perfused into the lumen of the cortical collecting tubule, two staining patterns emerged—bright surface staining with sharp edges (Fig. 1*h*), crescentic in perfused tubules, and more diffuse staining reminiscent of the endocytic pattern, with the nucleus observed as an unstained region (Fig. 1*h*). When the tubule was perfused with macromolecules labelled by a different fluorophore, only the cells with diffuse lectin staining endocytosed the macromolecules (compare Fig. 1*h* and *i*). Chronic acid loading did not increase the total number of intercalated cells per mm of tubule (control 129 ± 5 , $n = 15$ tubules; acid-loaded 137 ± 6 , $n = 10$ tubules, $P > 0.5$). Rather, the increase in the number of H^+ -secreting cells was accompanied by an equivalent decrease in the number of HCO_3^- -secreting cells (Table 2). Further, if the 10-fold increase in the number of H^+ -secreting cells seen after

Table 2 Effects of acid loading on isolated perfused cortical collecting tubules in rabbits

	$J_{\text{HCO}_3^-}$ ($\mu\text{mol min}^{-1}$ mm^{-1})	Potential difference (mV)	H^+ - secreting cells	HCO_3^- - secreting cells
Control	-1.6 ± 0.5 (4)	-13 ± 3 (4)	7 ± 2 (15)	122 ± 4 (15)
Acid-loaded	5.4 ± 0.8 (7)†	-5 ± 2 (7)*	74 ± 10 (10)†	63 ± 11 (10)†

Tubules were from control rabbits or from rabbits given NH_4Cl (10–20 mequiv per kg per day by gavage) for 5 ± 1 days. Cell counts for luminal endocytosis (10 control and 9 acid-loaded rabbits) were obtained by perfusing each tubule for 10 min at 38 °C with 0.5–1 mg ml⁻¹ fluorescein-dextran³ dissolved in a bicarbonate-free solution (290 ± 1 mosmol per kg) containing (in mM): NaCl (139), K_2HPO_4 (2.5), CaCl_2 (2.0), MgSO_4 (1.2), glucose (5.5), L-alanine (6.0), Na-lactate (4.0) and $\text{Na}_3\text{-citrate}$ (1.0); pH was 7.4. The chamber was then cooled to 9–12 °C for 5 min, after which rhodamine-peanut agglutinin was perfused in the lumen for 2–3 min. The solution was washed out with cold fluorophore-free solution and the tubule was warmed to room temperature for study by epifluorescence microscopy. At least three 250- μm segments of each cortical collecting tubule were examined for numbers of green, fluorescein-positive or luminally endocytosing cells, and for red or lectin-binding cells. Despite the cooling of the specimen chamber, it was not always possible to rule out endocytosis of peanut lectin by the acid-secreting cells, which could lead to an overestimate of the number of the HCO_3^- -secreting cells. The total number of cells per mm was determined by counting nuclei in duplicate at ×1,000 over three to six 50- μm lengths of each cortical collecting tubule that was fixed according to Fig. 1j legend. The total number of cells per mm of tubule was similar in both groups: control, 561 ± 8 ; acid-loaded, 543 ± 11 ($P > 0.2$). Bicarbonate transport was measured in a second perfused cortical collecting duct isolated from the same kidney slice. Transepithelial voltage (PD) was measured using the luminal pipette as an electrode and the bath solution as a reference³. The three to five measurements of potential difference were made at the time of each collection of tubular fluid for determination of bicarbonate concentration. Both luminal and bathing solutions were identical (290 ± 1 mosmol per kg) and contained (in mM): NaCl (114), NaHCO_3 (25), K_2HPO_4 (2.5), CaCl_2 (2.0), MgSO_4 (1.2), glucose (5.5), L-alanine (6.0), Na-lactate (4.0) and $\text{Na}_3\text{-citrate}$ (1.0); they were gassed with 95% O_2 /5% CO_2 to yield a pH of 7.4. Bicarbonate transport was measured by microcalorimetry as described in Table 1 legend. Ammonium chloride administration caused a chronic acidosis in most rabbits, the mean total CO_2 in the plasma being 22 ± 2 mM compared with 28 ± 1 mM in controls. Moreover, the urine pH was lower: 6.1 ± 0.4 in ammonium chloride-loaded rabbits and 8.5 ± 1.4 in control animals. Note that the administration of ammonium chloride did not necessarily cause a chronic acidosis in all animals, because the kidneys were markedly stimulated to increase acid excretion. Some animals showed normal levels of total CO_2 , yet the tubules from the animals absorbed bicarbonate at high rates and the number of cells that could endocytose fluorescent dextran from the lumen was markedly higher than in control animals.

* $P < 0.05$; † $P < 0.01$ (acid-loaded versus control). Mean ± s.e. (number of tubules).

a few days of acid loading was caused by cellular proliferation, mitotic figures should be apparent. We saw such figures (using the dye Hoechst 33258) only in the ampulla of the collecting duct of newborn rabbits (Fig. 1j). There was only one mitotic figure in 27 tubules from 5 control animals (15,150 cells) and none in 70 tubules from 10 acid-loaded animals (38,170 cells). To examine the possibility that mitosis occurred early in acid loading, we studied eight tubules (4,280 cells) removed from animals after 1 day of acid loading, but found no mitotic figures even though the number of H^+ -secreting cells had increased three to fivefold. Thus, the observed increase in the number of H^+ -secreting cells is caused by a reversal of the functional polarity of the HCO_3^- -secreting cell, and this reversal accounts for the conversion of the tubule from HCO_3^- secretion to H^+ secretion.

The conversion of HCO_3^- -secreting cells to acid-secreting cells requires not only relocation or activation/inactivation of the transporting elements (the proton ATPase, the parallel Cl^- channel and the $\text{Cl}^-/\text{HCO}_3^-$ exchanger) at the appropriate pole, but also remodelling of the cytoplasm to allow exocytosis and endocytosis at the apical pole of the cell and to inhibit them at the basolateral domain (Fig. 2). Such responses must also be stable for at least 24 h.

G.J.S. was supported by USPHS grant HD 13232 and by New York State Health Research Council grant 15-004, and is an Established Investigator of the American Heart Association; funds were contributed in part by the New York chapter. Q.A. was supported by USPHS grant AM 20999 and by a grant-in-aid from the New York Heart Association.

Received 22 May; accepted 21 August 1985.

- Steinmetz, P. R. *Physiol. Rev.* **54**, 890–956 (1974).
- Gluck, S., Cannon, C. & Al-Awqati, Q. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4327–4331 (1982).
- Schwartz, G. J. & Al-Awqati, Q. *J. clin. Invest.* **75**, 1638–1644 (1985).
- Cannon, C., van Adelsberg, J., Kelly, S. & Al-Awqati, Q. *Nature* **314**, 443–446 (1985).
- Le Furgey, A. & Tisher, C. C. *Am. J. Anat.* **155**, 111–123 (1979).
- Stone, D. K., Xie, X.-S. & Racker, E. *J. biol. Chem.* **258**, 4059–4062 (1983).
- Galloway, C. J., Dean, G. E., Marsh, M., Rudnick, G. & Mellman, I. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3334–3338 (1983).
- Forgac, M., Cantley, L., Wiedemann, B., Altstiel, L. & Branton, D. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1300–1303 (1983).
- Le Hir, M., Kaissling, B., Koeppen, B. M. & Wade, J. B. *Am. J. Physiol.* **242**, C117–C120 (1982).
- Schwartz, G. J. & Evan, A. P. *Am. J. Physiol.* **245**, F382–F390 (1983).
- Itoh, K. *et al. Brain Res.* **175**, 341–346 (1979).

Upstream sequences modulate the internal promoter of the human 7SL RNA gene

Elisabetta Ullu* & Alan M. Weiner

Department of Molecular Biophysics and Biochemistry,
Yale University, PO Box 3333, New Haven, Connecticut 06510, USA

The human genome is rich in sequences which are structurally related to the 7SL RNA component of the signal recognition particle¹. The 7SL DNA sequence family consists of four 7SL genes², 500 7SL pseudogenes² (which are truncated at one or both ends of the 7SL sequence) and 500,000 *Alu* sequences^{3–5}. Both 7SL genes² and *Alu* elements^{6–8} are transcribed by RNA polymerase III, and we show here that the internal 7SL promoter lies within the *Alu*-like part of the 7SL gene. Why then does RNA polymerase III transcribe the few 7SL genes so efficiently, while transcripts from the far more abundant *Alu* elements are not readily detectable^{9–11}? We find that a human 7SL gene and a synthetic *Alu* sequence derived from it are expressed 50–100-fold more efficiently *in vitro* than either a representative *Alu* element or two 7SL pseudogenes. 5' Deletion and insertion mutants of the 7SL gene demonstrate that, in conjunction with the internal promoter, the first 37 nucleotides upstream from the transcription start site are essential for efficient and accurate initiation *in vitro*. We suggest that the genomic sequences upstream from most *Alu* elements and 7SL pseudogenes do not contain this element, and consequently that only a small subset of such sequences can be transcribed *in vivo*. This may help to explain the homogeneity of the *Alu* family within each mammalian genome, as well as the species-specific differences between mammalian *Alu* families.

To determine which sequences in the human 7SL gene are important for expression, we compared the *in vitro* transcription of one of the four human 7SL genes, 7L30.1 (ref. 2) (Fig. 1A, lane 30), two 3'-truncated 7SL pseudogenes, 7L28 and 7LEM1 (ref. 2) (lanes 28 and EM1), and a pBR322 derivative of the human 7SL cDNA clone p7L1 (ref. 12), called p7L1-bis (lane 1; see Fig. 1B for the structures of the clones). Although all the 7SL clones were transcribed, the wild-type 7SL gene was 50–100 times more efficient than the other templates. The gene-pseudogene fusions (lanes 30–28 and 28–30) show that the sequences required for efficient transcription lie predominantly to the 5' side of the *Sma*I site in the 7SL gene (Fig. 1B). Transcription of the 7SL cDNA clone p7L1-bis proves that the 7SL coding region contains an internal promoter.

We also compared the transcription of a cloned *Alu* element with that of the 7SL gene (Fig. 1C). For this analysis we chose the *Alu* sequence located upstream from the human γ -globin gene, which is transcribed *in vitro* to yield a major transcript of ~575 nucleotides⁶. Since full activity of the 7SL promoter² requires an unusually low concentration of Mg^{2+} , the *Alu* and 7SL gene templates were assayed over a range of Mg^{2+} concentrations. The optimum for the 7SL gene was 1 mM Mg^{2+} ,

* Present address: Department of Internal Medicine, Yale University, PO Box 3333, New Haven, Connecticut 06510, USA.

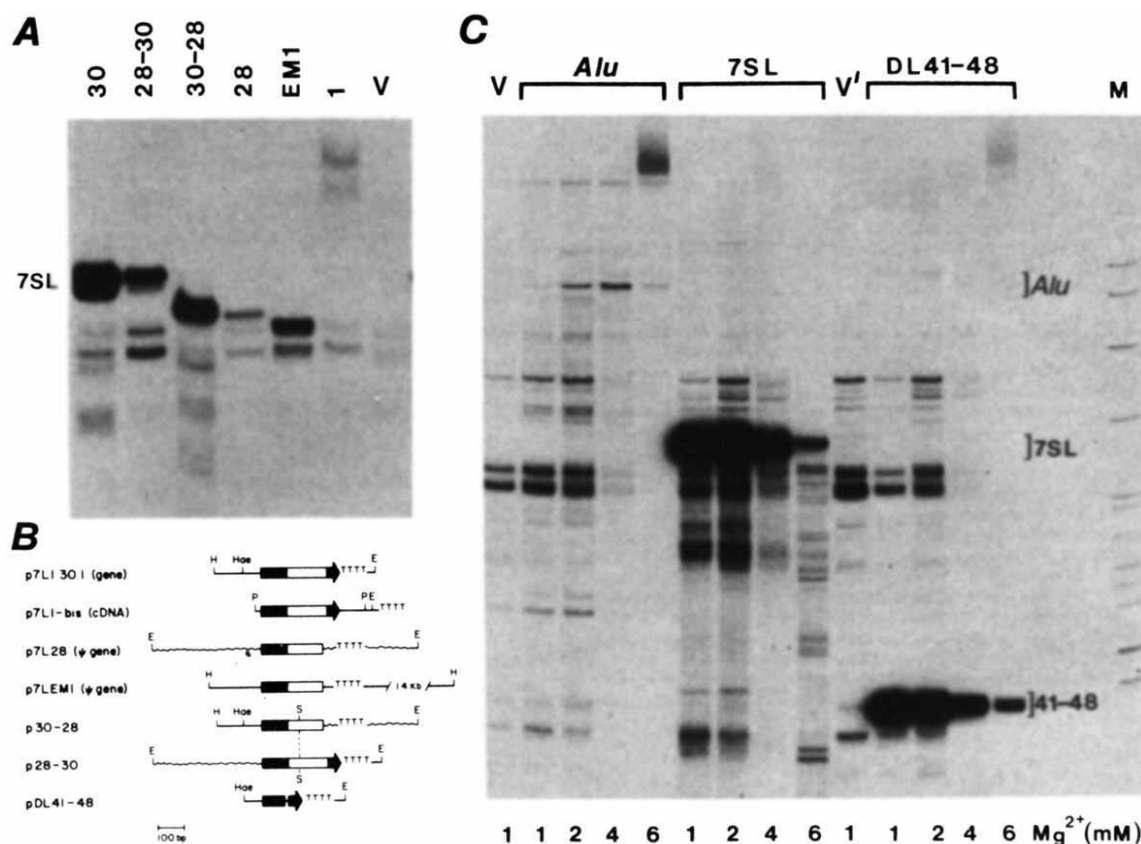


Fig. 1 **A**, The products of *in vitro* transcription of gene 7L30.1² (lane 30, 7SL RNA-300 nucleotides; nt), 3'-truncated pseudogenes 7L28 (lane 28, 7L28 RNA ~275 nt) and 7LEM1 (lane EM1, 7LEM1 RNA 265 nt)², 7L1-bis cDNA (lane 1, 7L1 RNA ~440 nt), recombinants 7L30-28 (lane 30-28) and 7L28-30 (lane 28-30) and pUC13 vector DNA (lane V). Transcription of 7SL templates in a HeLa cell nuclear extract³¹, fractionation and purification of the transcripts were carried out as described previously². To quantitate RNA synthesis, labelled RNA bands were detected by autoradiography, excised from the gel and Cherenkov radiation was counted. In different experiments the yield of ³²P radioactivity per band was: 7L30.1 RNA, 2.5–4 × 10⁵ c.p.m.; 7L28 RNA, 2–4 × 10⁵ c.p.m.; 7LEM1 RNA, 5–10 × 10⁵ c.p.m.; 7L1-bis RNA 2–4 × 10⁵ c.p.m.; 7L30-28 RNA, 1.5–2 × 10⁵ c.p.m.; 7L28-30 RNA 5–10 × 10⁵ c.p.m. As assayed by T1 RNAase fingerprinting, the major transcripts derived from all templates initiated at or near the normal start site² and terminated at a run of four or more T residues in the template DNA (see **B**; data not shown). **B**, Schematic representation of the 7SL clones used for *in vitro* transcription. p7L30.1² contains human 7SL gene 7L30.1 including the entire coding region plus ~180 base pairs (bp) of 5'-flanking and 30 bp of 3'-flanking DNA in pUC13. p7L28 and p7LEM1 are pUC13 plasmid subclones derived from 7SL pseudogenes 7L28 and 7LEM1². Both pseudogenes are colinear with the 7SL coding sequence up to position 235, but their genomic context is different. p7L1-bis contains a complete cDNA copy of human 7SL RNA from clone p7L1¹² inserted into pBR322 so that the T₅ oligonucleotide at vector position 4,324 is immediately 3' to the 7SL coding region. Plasmid subclones p7L30.1 and p7L28 were recombined using the unique *Sma*I site at position 143 in the 7SL coding region. The DNA fragment corresponding to the 5' half of the gene was ligated to the 3' fragment of the 7L28 pseudogene to create p7L30-28, and the 5' half of the pseudogene was ligated to the 3' half of the gene to create p7L28-30. Plasmid pDL41-48 is a recombinant derived from p7L30.1-5'-66 (see below) in which the nucleotides between positions 85 and 248 have been deleted and substituted with an 8-nt *Xho*I linker. The arrows represent the various 7SL DNAs and indicate the direction of transcription; the shaded areas represent the *Alu* sequence of 7SL DNA^{4,12}; H, *Hind*III; E, *Eco*RI; S, the *Sma* site used to recombine p7L30.1 and p7L28; P, *Pst*I; Hae, *Hae*III; TTTT, indicates the position of the transcription terminator³². **C**, Products of *in vitro* transcription of the 7SL gene 7L30.1 (lanes 7SL), clone pBS2⁶ (lanes *Alu*) which contains the *Alu* sequence upstream from the human γ -globin gene, and of clone pDL41-48 (lanes 41-48) as a function of varying Mg²⁺ concentration. The amount of exogenous Mg²⁺ added is indicated below each lane. V, pBR322 transcripts; V', pUC13 transcripts; M, end-labelled *Hpa*II DNA fragments of pBR322 as relative molecular mass markers. The yield of transcripts from clone pBS2 was comparable to that from pseudogene 7L28. The human genomic *Alu* clone pEB77 (ref. 33) and the monkey *Alu* clone p7.01 (ref. 34) gave a similar result.

while the *Alu* sequence was transcribed best at 4 mM Mg²⁺. Even when the transcription activities of the two templates were compared at their respective Mg²⁺ optima, the 7SL gene was at least 100 times more active than the genomic *Alu* clone. A similar result was obtained with two additional genomic *Alu* sequences (see Fig. 1C legend). Taken together, these results suggest that either the 5'-flanking DNA or the central non-*Alu* portion of the 7SL coding region (the S sequence¹²) might be responsible for the high level of 7SL gene transcription.

To test whether sequences in the non-*Alu* portion of the 7SL coding region are required for efficient transcription, we deleted the S sequence from the 7SL gene. This synthetic *Alu* sequence (pDL41-48, Fig. 1B) produced the expected shorter transcript

of ~140 nucleotides (Fig. 1C, lanes DL41-48), was as active as the 7SL gene itself, and exhibited the same Mg²⁺ optimum. We conclude that the 5'-flanking sequence of gene 7L30.1 is likely to be responsible for its high expression, and that the internal control region must reside within the *Alu*-like parts of 7SL DNA because the S sequence is not required for transcription *in vitro*.

Next, we constructed a series of 7SL genes with progressive deletions of 5'-flanking DNA (Fig. 2A). Deletion of sequences downstream of nucleotide -37 had two effects on *in vitro* transcription (Fig. 2B). First, mutant 7SL genes with less than 37 base pairs of 5'-flanking sequence displayed a significant reduction in activity. For instance, the 5'-2 mutant, which contains only the 7SL coding sequence plus 30 nucleotides of 3'-flanking

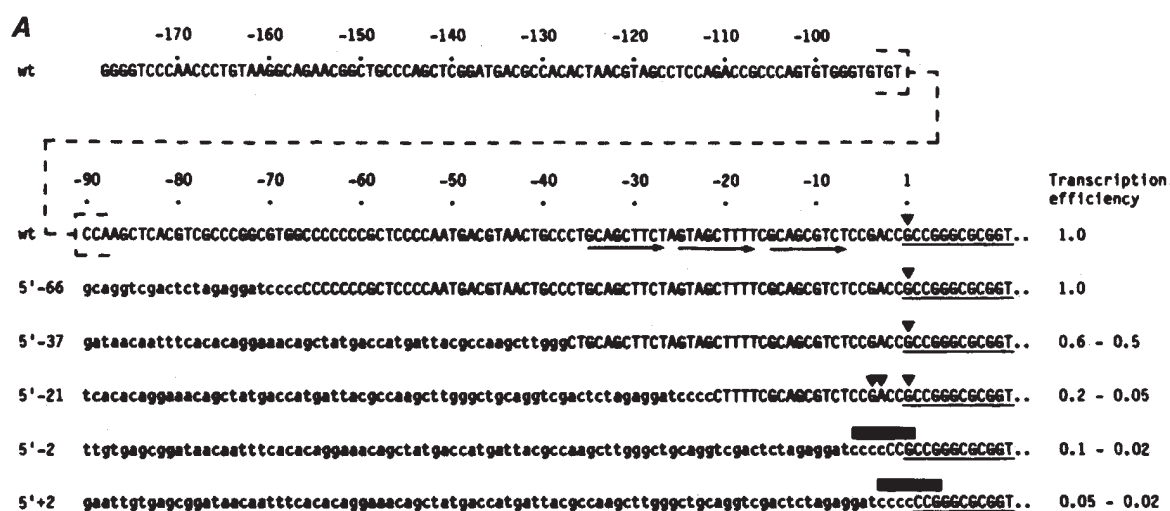


Fig. 2 A, Nucleotide sequences and transcription efficiencies of the 5' deletion derivatives of the 7SL gene 7L30.1 (wt). Progressive deletions of the 5'-flanking DNA of gene 7L30.1 were generated either by digestion with restriction enzymes or with *Bal31* nuclease. The 5' deletion derivatives are named according to the position of the deletion end point. Structures of the deleted templates were verified by DNA sequence analysis using the chemical degradation method³⁵. The 5'-37 recombinant was constructed by digestion of p7L30.1 with *Pst*I plus *Eco*RI, followed by insertion of the resulting fragment at the corresponding sites in pUC13 DNA. The other 5' deletions were inserted between the *Sma*I and *Eco*RI sites of the same plasmid vector. The beginning of the 7SL RNA coding sequence is underlined. Upper case letters: natural 5'-flanking DNA; lower case letters: pUC13 vector DNA. Arrowheads: transcription initiation sites mapped by T1 fingerprint and 5' end-group analysis. Solid bar: presumptive initiation sites; we did not determine the precise initiation sites of 5'-2 and 5'+2 RNAs, but their 5' ends mapped within the cluster of C residues abutting the 7SL coding sequence. Arrows below the wild-type sequence indicate a set of three 9-bp direct repeats. The transcription efficiencies represent both the highest and the lowest obtained using several different preparations of HeLa cell nuclear extract. Note that a sequence closely resembling the A block consensus sequence of transfer RNA promoters³⁶ is located between positions 5 and 13 of the 7SL gene². Indeed, the first 15 nucleotides upstream from the 7SL coding sequence are essential for expression *in vitro* (E.U., unpublished observations). **B**, *In vitro* transcription products of the 7SL gene 7L30.1 (wt) and of the 5' deletion derivatives. Numbers above each lane indicate the positions of the deletion end points. V, the products of transcription of pUC13 vector DNA. The wet gel was exposed to a Kodak X-AR5 film for 5 min. The faint band migrating faster than 7SL RNA is a 7SL transcript which is truncated at the 3' end (data not shown) perhaps as result of premature termination or a 3' exonuclease activity. **C**, As in **B** except that the gel was exposed to a Cronex X-ray film for 2 min. **D**, *In vitro* transcription products of gene 7L30.1 (a), the 5'-2 deletion mutant (b) and a pool of 13 independent 5'-2 recombinants (c). *Sau*3A λ DNA fragments were inserted into the *Bam*HI site at position -10 of the 5'-2 clone; 13 independent recombinants were chosen carrying different *Sau*3A fragments, equal amounts of the various DNAs were pooled, and their cumulative transcription activity was determined. **E**, Transcription of the 7SL gene (lane 30), the 5'-21 (lane 21) and the 5'-2 (lane 2) deletion mutants in injected oocytes. *Xenopus laevis* oocytes were injected with ~20 nl of a solution containing plasmid DNA (400 μ g ml⁻¹) and radioactive nucleotide precursors. *Xenopus borealis* 5S maxigene DNA¹⁵ (10 μ g ml⁻¹) was co-injected as an internal control for the efficiency of injection. For each template, a set of 20 oocytes was injected and incubated for 20-24 h at 20 °C. 5S indicates the products of transcription of the *X. borealis* 5S RNA maxigene.

DNA, was 10-50 times less active than the wild-type gene. The upstream vector sequences do not contain an inhibitor of transcription, because replacement of vector DNA by a random collection of bacteriophage λ DNA fragments had no effect (Fig. 2D). Second, a good proportion of the transcripts produced from these mutant 7SL genes were larger than 7SL RNA (Fig. 2C). As judged by T1 fingerprinting and 5'-end-group analysis, approximately 10% of the transcripts from deletion mutants 5'-21, 5'-2 and 5'+2 were initiated correctly, while the remainder had 5' ends mapping a few nucleotides upstream from the G residue at position 1 (see Fig. 2A; data not shown). The failure of these 7SL deletion mutants to accumulate significant levels of 7SL RNA must reflect a decrease in the efficiency of transcription, since pulse-chase experiments *in vitro* did not reveal any difference in the turnover rate of the 5'-2 RNA as compared with authentic 7SL RNA (data not shown).

To determine whether the efficiency of transcription depends on the distance between the 5'-flanking sequences and the 7SL coding region, we constructed a second set of mutant 7SL genes

with 9, 11 and 13 extra nucleotides inserted at position -6 of the 7L 5'-66 template (Fig. 3A). Relative to the parent template (Fig. 3B), these mutations caused a ~10-fold reduction in transcription activity, and led to aberrant initiation upstream, as assayed by primer extension (Fig. 3C). Further increases in the length of the linker DNA inserted at position -6 did not significantly change the pattern of initiation observed with mutant DL+13 (data not shown). We conclude that the sequences upstream from the 7SL gene, in conjunction with the internal control region, provide information for positioning the correct start site of transcription; moreover, optimal transcription requires that the distance of the upstream DNA from the internal promoter is not altered.

Finally, the 5'-flanking sequence of the 7SL gene increases the efficiency of transcription from the 7SL internal promoter *in vivo* as well as *in vitro*. After injection into the nucleus of *Xenopus* oocytes, the 5'-21 and 5'-2 mutant templates (Fig. 2E) were transcribed 5-10-fold less efficiently than the wild-type gene, even though this is a heterologous system. In addition, the 5'-21 and 5'-2 transcripts were longer than 7SL

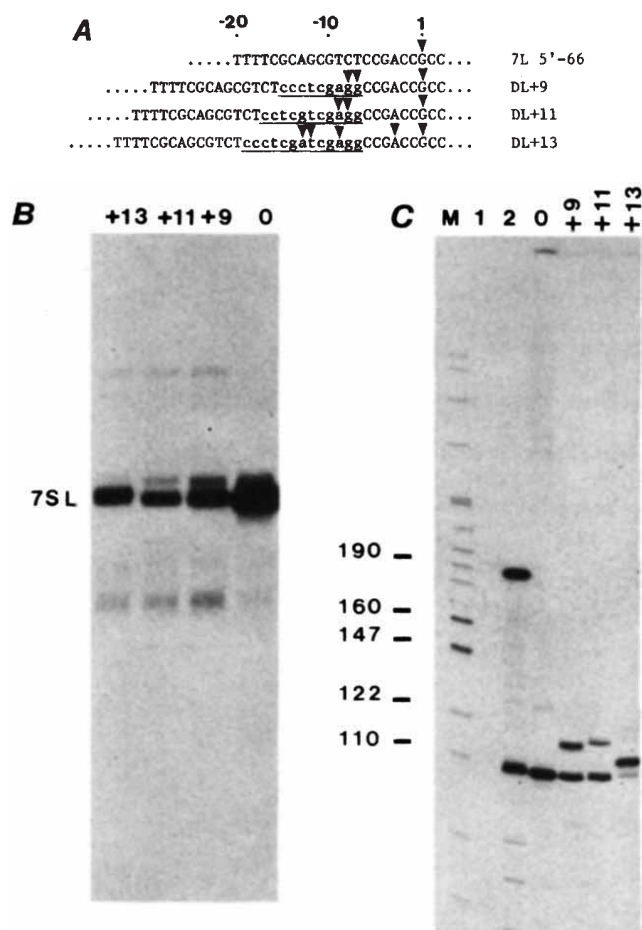


Fig. 3 **A**, Nucleotide sequences of mutant 7SL genes *DL+9*, *DL+11* and *DL+13* and position of the initiation sites mapped by primer extension (▼). The linker sequences are underlined. The *DL+9* recombinant was constructed by inserting an *Xho*I linker at position -6 of 7L 5'-66 DNA, after digestion³⁷ with DNase I in the presence of Mn^{2+} . *DL+11* and *DL+13* were obtained by filling the ends of *Xho*I-digested *DL+9* DNA with Klenow polymerase and deoxynucleotide triphosphates. **B**, Products of *in vitro* transcription of 7SL mutants *DL+9* (lane +9), *DL+11* (lane +11) and *DL+13* (lane +13) and of 7L 5'-66 (lane 0). **C**, Primer extension analysis of 7SL RNA and of *DL+9*, *DL+11* and *DL+13* RNAs synthesized *in vitro*. The following RNAs were used as templates: lane 1, endogenous RNAs after micrococcal nuclease digestion; lane 2, endogenous RNAs before micrococcal nuclease digestion; lane 0, 7SL RNA synthesized from 7L 5'-66 DNA; lane +9, *DL+9* RNAs; lane +11, *DL+11* RNAs; lane +13, *DL+13* RNAs. M, end-labelled *Hpa*II fragments of pBR322 DNA.

Methods. The 5' ends of the transcripts were mapped by primer extension using an oligonucleotide complementary to positions 96-110 of 7SL RNA. The endogenous RNAs of the transcription extract³¹ gave two major extension products of 110 and 190 nt (lane 2); the 110-nt cDNA corresponds to 7SL RNA, while the origin of the longer product is unknown. The extract was therefore digested with micrococcal nuclease before transcription (lane 1) to remove these endogenous RNAs. Micrococcal nuclease (5,000 U ml⁻¹) was added to HeLa cell nuclear extract containing 1 mM CaCl₂ and the digestion was carried out for 15 min at 30 °C. Nuclease digestion was stopped by addition of 2 mM EGTA; transcription reactions contained 500 µM of each of the four ribonucleotide triphosphates. Nucleic acids were purified by phenol extraction and, after ethanol precipitation, annealed to a 10-fold molar excess of the 5'-end-labelled synthetic oligonucleotide. Reverse transcription was carried out as described by Hernandez and Keller³⁸. cDNA products were fractionated through a 6% polyacrylamide gel containing 7 M urea.

RNA and identical in structure to the corresponding transcripts synthesized *in vitro*.

Our results have evolutionary implications. *Alu* elements are thought to transpose to new chromosomal sites using an RNA polymerase III transcript as an intermediate^{13,14}. The 5' end of the transposed *Alu* information is therefore defined by the initiation site for transcription of the parental *Alu* element, and each newly transposed *Alu* element acquires a new 5'-flanking sequence. Deletions and substitutions of the 5'-flanking sequences in a variety of genes transcribed by RNA polymerase III genes are known to affect the accuracy of initiation¹⁵⁻¹⁷ and to modulate either positively¹⁸⁻²² or negatively²³⁻²⁵ the efficiency of transcription. We have shown here that alterations in the 5'-flanking DNA of the human 7SL gene markedly decrease the efficiency of transcription and lead to the appearance of heterogeneous start sites. We speculate that the 5'-flanking sequences of most newly inserted *Alu* elements are unlikely to be compatible with maximal transcription and that only a small subset of mammalian *Alu* family members account for most *Alu* transcription *in vivo*. This subset of *Alu* elements ('founder elements') would produce most of the transcripts that serve as intermediates for transposition to new chromosomal sites. Mutations occurring in these founder elements when the copy number of *Alu* sequences was low at the time of the mammalian radiation, would lead to the observed species-specific differences between mammalian *Alu* sequences²⁶⁻²⁹, as the copy number of *Alu* sequences increased in each species³⁰.

We thank Doug Black for the 7SL oligonucleotide, Manny Ares and Chang-Yee Yuo for teaching us how to inject oocytes and Manny Ares, Nouria Hernandez, Ellen Gottlieb and Chris Tschudi for the critical reading of the manuscript. E.U. was the recipient of a Grant-in-Aid from the Sigma Delta Epsilon, Graduate Women in Science, Inc. This work was supported by NIH grants GM 31073 and 31335.

Received 8 July; accepted 26 September 1985.

- Walter, P. & Bobel, G. *Nature* **299**, 691-698 (1982).
- Ullu, E. & Weiner, A. M. *EMBO J.* **3**, 3303-3310 (1984).
- Houck, C. M., Rinehart, F. P. & Schmid, C. W. *J. molec. Biol.* **132**, 289-306 (1979).
- Ullu, E. & Tschudi, C. *Nature* **312**, 171-172 (1984).
- Gundelfinger, E. D., Di Carlo, M., Zopf, D. & Melli, M. L. *EMBO J.* **3**, 2325-2332 (1984).
- Duncan, C. H., Jagadeeswaran, P., Wang, R. R. C. & Weissman, S. H. *Gene* **13**, 185-196 (1981).
- Paoletti, G., Lucero, M. A., Murphy, M. H. & Baralle, F. *EMBO J.* **2**, 691-696 (1983).
- Perez-Stable, C., Ayres, T. M. & Shen, C. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5291-5295 (1984).
- Haynes, S. R. & Jelinek, W. R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6130-6134 (1981).
- Young, P. R., Scott, R. W., Homer, D. H. & Tilghman, S. M. *Nucleic Acids Res.* **10**, 3099-3116 (1982).
- Allan, M. & Paul, J. *Nucleic Acids Res.* **12**, 1193-1200 (1984).
- Ullu, E., Murphy, S. & Melli, M. *Cell* **29**, 195-202 (1982).
- Van Arsdell, S. W. *et al. Cell* **26**, 11-17 (1981).
- Jagadeeswaran, P., Forget, B. G. & Weissman, S. M. *Cell* **26**, 141-142 (1981).
- Sakonju, S., Bogenhagen, D. F. & Brown, D. D. *Cell* **19**, 13-25 (1980).
- Hofstetter, H., Kressmann, A. & Birnstiel, M. L. *Cell* **24**, 573-585 (1981).
- Fowlkes, D. M. & Shenk, T. *Cell* **22**, 405-413 (1980).
- Larson, D., Bradford-Wilcox, J., Young, L. S. & Sprague, K. U. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3416-3420 (1983).
- Raymond, G. J. & Johnson, J. D. *Nucleic Acids Res.* **11**, 5969-5988 (1983).
- Morton, D. G. & Sprague, K. U. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5519-5522 (1984).
- Shaw, K. J. & Olson, M. V. *Molec. cell. Biol.* **4**, 657-665 (1984).
- Schaack, J. *et al. J. biol. Chem.* **259**, 1461-1467 (1984).
- De Franco, D., Schmidt, O. & Söll, D. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3365-3368 (1980).
- Dingermann, T., Burke, D. J., Sharp, S., Shaack, J. & Söll, D. *J. biol. Chem.* **257**, 14738-14744 (1982).
- Hipskind, R. A. & Clarkson, S. G. *Cell* **34**, 881-890 (1983).
- Krayev, A. S. *et al. Nucleic Acids Res.* **8**, 1201-1215 (1980).
- Haynes, S. R., Toomey, T. P., Leinwand, L. & Jelinek, W. R. *Molec. cell. Biol.* **1**, 573-583 (1981).
- Deininger, P. L., Jolly, D. J., Rubin, C. M., Friedmann, T. & Schmid, C. W. *J. molec. Biol.* **151**, 17-33 (1981).
- Daniels, G. R., Fox, G. M., Loewenstein, D., Schmid, C. W. & Deininger, P. L. *Nucleic Acids Res.* **11**, 7569-7593 (1983).
- Weiner, A. M., Deininger, P. L. & Efstradiatis, A. A. *Rev. Biochem.* **55** (in the press).
- Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. *Nucleic Acids Res.* **11**, 1475-1489 (1983).
- Bogenhagen, D. F. & Brown, D. D. *Cell* **24**, 261-270 (1981).
- Ullu, E. & Melli, M. L. *Nucleic Acids Res.* **10**, 2209-2223 (1982).
- Grimaldi, G., Queen, C. & Singer, M. *Nucleic Acids Res.* **21**, 5553-5567 (1981).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
- Ciliberto, G., Rauegi, G., Costanzo, F., Dente, L. & Cortese, R. *Cell* **32**, 725-733 (1983).
- Heffron, F., So, M. & McCarthy, B. J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6012-6016 (1978).
- Hernandez, N. & Keller, W. *Cell* **35**, 89-99 (1983).

Tailoring the pH dependence of enzyme catalysis using protein engineering

Paul G. Thomas, Alan J. Russell & Alan R. Fersht

Department of Chemistry, Imperial College of Science and Technology, London SW7 2AY, UK

One of the goals of protein engineering is to tailor the pH dependence of enzyme catalysis to optimize activity in industrial processes. Chemical modification studies have shown that the pH dependence of catalysis by serine proteases alters with changes in overall surface charge¹, which suggests that a possible general method of modifying pH dependence is to alter the electrostatic environment of the active site by protein engineering and so change the pK_a values of ionic catalytic groups. Electrostatic effects are of considerable importance in enzyme catalysis and are thought to play a major part in stabilizing charged transition states²⁻⁵. However, it is extremely difficult to calculate electrostatic effects in proteins because of the microheterogeneity of the dielectric constant. In the present study we show how the change of just one surface charge which is 14–15 Å from the active site of subtilisin has a significant effect on the pH dependence of the enzyme and that the magnitude of the effect changes with ionic strength in a manner qualitatively consistent with electrostatic theory. Such experiments should provide the basis for refining theoretical calculations of electrostatic effects in catalysis.

The subtilisins are a family of extracellular serine proteases (also known as alkaline proteases) produced by species of bacilli prior to sporulation⁶. His 64 at the active site of the enzyme acts as a general base during catalysis, accepting a proton from the nucleophilic residue Ser 221 as it forms a bond with the substrate carbonyl carbon. The enzyme is active at alkaline pH when His 64 is unprotonated, and catalytic activity varies with pH following the ionization of this residue. The mutation we have chosen for our investigation of the magnitudes of electrostatic effects is Asp→Ser 99 in subtilisin from *Bacillus amylo-liquefaciens* (BPN' or Novo). Crystallographic data show that this surface aspartate is some 14–15 Å from the imidazole of His 64 (ref. 7). (Note that because of a sequencing error, there is an alanine at position 99 in the early crystallographic models. We built in an aspartate at this position using molecular graphics. The derived distances have been substantiated by a new refined structure (R. Bott, personal communication)—the His 64 C_γ–Asp C_γ distance is 13 Å in the refined map, the same as in our modelling.) Asp 99 is not conserved in highly homologous subtilisins from other organisms: there is a serine at this position in the subtilisins from *Bacillus subtilis* DY⁸ and *Bacillus licheniformis*⁹ and threonine in subtilisins from *B. subtilis*¹⁰ and *Bacillus amylosacchariticus*¹¹. It is thus expected *a priori* that the mutation Asp→Ser 99 should not have a significant effect on either the structure or catalytic properties of the enzyme other than by electrostatic effects. Removal of the negative charge of Asp 99 should destabilize the low-pH, positively charged form of His 64 and so lower its pK_a from the normal value of 7 by an unknown amount.

To test the predictions and quantify the effects, we have cloned the structural gene encoding subtilisin BPN' from *B. amylo-liquefaciens* in the vector pUB110 (refs 12, 13). The structure of pPT30, which harbours the subtilisin gene on a 3.4-kilobase (kb) *EcoRI* fragment, is shown in Fig. 1a; its construction will be described elsewhere. The DNA sequence of the structural gene was determined by dideoxy methods¹⁴ using a series of primers (J. Brannigan, unpublished data) and is identical to that published elsewhere^{15,16}. Wild-type subtilisin BPN' is expressed and secreted at high levels by *B. subtilis* DB104 transformed with pPT30 (Fig. 1b). The strain DB104, given by Dr R. Doi, is deficient in extracellular alkaline and neutral proteases¹⁷. Subtilisin(Asp→Ser 99), prepared by oligodeoxynucleotide

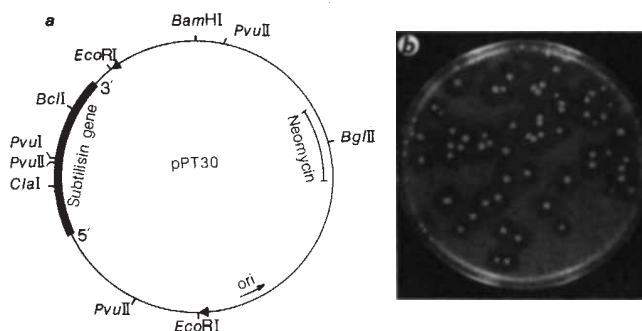


Fig. 1 a, Restriction map of the subtilisin expression vector pPT30. The subtilisin gene is contained within a 3.4-kb *EcoRI* fragment (solid block). Sequences derived from pUB110 (4.5 kb) are delineated by the arrowheads. b, Secreted subtilisin gives a characteristic clear halo around colonies of *B. subtilis* DB104 harbouring pPT30 after overnight selection on L-agar containing 1% skimmed milk plus 25 µg ml⁻¹ kanamycin.

Methods. For mutagenesis, the gene was recloned in M13mp9. M13aprWT was constructed such that the coding strand of the subtilisin gene was ligated to the (+) strand of the viral DNA. M13aprWT was propagated in *Escherichia coli* TG2, a *recA*⁻ derivative of JM101 (M. Biggin and T. Gibson, unpublished results) to overcome problems of DNA instability encountered when JM101 was the host. The synthetic primer (5'-CGGAACCGCTAG-CACG3') contained a double mismatch with the wild-type gene (indicated by asterisks) and was designed to introduce an acceptable Ser codon at position 99. Oligodeoxynucleotide-directed mutagenesis was performed as described elsewhere^{20,21}, except that the mismatch repair-deficient strain BMH71-18mutL²² was the recipient in transformations. TG2 provided the lawn to isolate phage from the mutator background. Hybridization screening using the mutagenic primer as a probe showed that 20% of the plaques contained mutant phage. The mutation was confirmed by DNA sequencing after plaque purification. Subtilisin(Asp→Ser 99) was expressed in DB104 on transformation of competent cells²³ with a plasmid constructed by ligating the mutant gene isolated from replicative form DNA into pUB110 in the same orientation as pPT30. Wild-type and mutant enzymes were purified from 36-h culture supernatants grown in L-broth containing kanamycin (25 µg ml⁻¹) and glucose (0.2%). The supernatant was concentrated using a Pellicon membrane filter (PTGC00005, Millipore) before dialysis against 0.01 M phosphate buffer pH 6.2 and purification on a CM-52 cellulose column as described elsewhere (ref. 24).

mutagenesis (Fig. 1), is also expressed at similar high levels in DB104.

The activity of subtilisin(Asp→Ser 99) on the synthetic substrate succinyl-L-alanyl-L-alanyl-L-prolyl-L-phenylalanyl *p*-nitroanilide is similar to that of wild-type enzyme at 25 °C in 0.1 M Tris-HCl buffer, pH 8.6: k_{cat} (the catalytic rate constant) for the wild-type enzyme is 57 s⁻¹ and K_m (Michaelis constant) is 0.15 mM, while the corresponding values for the mutants enzyme are 45 s⁻¹ and 0.13 mM. The wild-type enzyme has identical activity to that produced from an independently cloned gene (ref. 15 and D. A. Estell and J. A. Wells, personal communication). The pK_a values of the active sites of the free enzymes were determined by kinetics from the pH dependence of k_{cat}/K_m (Fig. 2, Table 1). At ionic strength 0.1, His 64 in the wild-type enzyme has a pK_a of 7.17 ± 0.02; this is decreased to a value of 6.88 ± 0.03 in the mutant, a shift of 0.29 ± 0.04 units. k_{cat}/K_m is decreased by 20% on mutation. Thus, Asp 99 affects both the ionization constant of the active site and also the charged transition state of the reaction.

As a control, it is important to examine the effect of ionic strength on the electrostatic interactions. The presence of high concentrations of ions should mask electrostatic interactions. At high ionic strengths, the observed pK_a s should tend to their intrinsic values in the absence of surface charge. According to the classical Debye-Hückel theory and the Poisson-Boltzmann

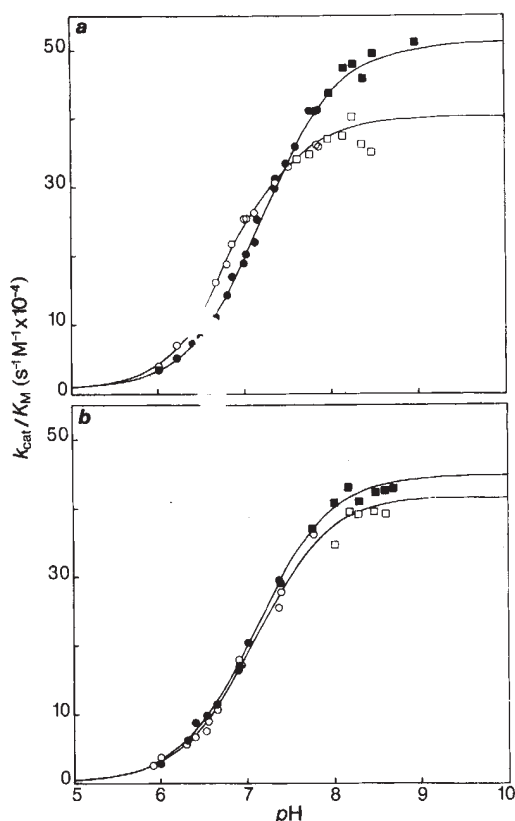


Fig. 2 pH dependence of k_{cat}/K_M for the hydrolysis of succinyl-L-alanyl-L-alanyl-L-prolyl-L-phenylalanyl *p*-nitroanilide by wild-type (solid symbols) and mutant Asp→Ser 99 (open symbols) subtilisin at 25 °C, ionic strength 0.1 (a) or 1.0 (b). Phosphate (circles) and Tris-HCl (squares) buffers of ionic strength 0.1 were used and the ionic strength adjusted by the addition of KCl. The reaction was initiated by adding 15 μl of enzyme (13 μM in 0.01 M phosphate buffer, pH 6.2) to 2 ml of buffered substrate (0.02 mM, $\ll K_M$) in a cuvette maintained at 25 °C in a Gilford 2600 spectrophotometer. The increase in absorbance at 412 nm on the release of *p*-nitroaniline was monitored and found to follow, with high accuracy, simple first-order kinetics. The value of k_{cat}/K_M at each pH value was determined from the first-order plots exactly as described for the hydrolysis of acetyl-L-tyrosine *p*-acetylanilide by chymotrypsin²⁵. The low concentration of substrate used in this method avoids product inhibition. The concentration of each enzyme was determined by active-site titration with *N*-trans-cinnamoyl imidazole²⁶. Wild-type enzyme was found to be 91% active and the mutant 96% active compared with concentrations measured by A_{280} ($E_{0.1\%} = 1.17$). Control experiments showed that the rate of reaction in Tris-HCl buffers was 15% lower than in phosphate buffers. The data for experiments in the presence of Tris-HCl have been corrected for this, but these data were not used in the analysis. The data for the experiments in phosphate buffer were fitted to a theoretical single ionization curve by nonlinear regression (R. J. Leatherbarrow, unpublished). The derived pK_a values are listed in Table 1.

equation, the electrostatic interaction between two ions separated by a distance r is reduced by a factor of $\exp(-r/r_D)$ where r_D , the Debye length, is inversely proportional to the square root of the ionic strength. At ionic strength 1.0, r_D is 3.1 Å and at ionic strength 0.1 it is 9.8 Å. Although the theory is only approximate at high ionic strength, electrostatic effects should be highly masked over a distance of 14–15 Å at ionic strength 1.0. Accordingly, we found (Table 1) that at this ionic strength, both the pK_a values of the active site and the catalytic activities of the wild-type and mutant enzymes converge ($pK_a = 7.11 \pm 0.02$ and 7.09 ± 0.02 ; $k_{\text{cat}}/K_M = (45 \pm 0.6) \times 10^4$ and $(42 \pm 1) \times 10^4 \text{ s}^{-1} \text{ M}^{-1}$ respectively). Thus, the mutation Asp→Ser 99 has essentially no effect on the catalytic properties of the enzyme when electrostatic effects are masked.

Table 1 The pH dependence of hydrolysis by wild-type and mutant subtilisin

Enzyme	pK_a of active site (from kinetics)	Limiting value of k_{cat}/K_M at high pH ($\text{s}^{-1} \text{ M}^{-1} \times 10^{-4}$)
Ionic strength 0.1		
Wild type	7.17 ± 0.02	48.9 ± 0.7
Asp→Ser 99	6.88 ± 0.03	40.8 ± 0.9
Ionic strength 1.0		
Wild type	7.11 ± 0.02	44.8 ± 0.6
Asp→Ser 99	7.09 ± 0.02	41.6 ± 0.9

The pK_a values were obtained from the pH dependence of k_{cat}/K_M on the hydrolysis of succinyl-L-alanyl-L-alanyl-L-prolyl-L-phenylalanyl *p*-nitroanilide at 25 °C. The pH dependence of k_{cat}/K_M gives the pK_a values of the free enzyme and free substrate. In this case, the only ionizing group on the substrate is the succinyl group and its pK_a is below the range of this study. Artefacts which often obscure the interpretation of the pH dependence of k_{cat} and K_M individually usually do not affect the pH dependence of k_{cat}/K_M (ref. 18). The pK_a of the wild-type enzyme is in excellent agreement with that derived from the hydrolysis of ester substrates in identical conditions (7.15 at ionic strength 0.1; ref. 19). Values are given $\pm$ s.e.

We have shown that modification of a single charge in the vicinity of the active site of an enzyme can have a significant effect on the pH dependence of the catalytic reaction. Such effects should be larger for charges which are closer to ionic catalytic residues in the enzyme or closer to charges which develop in the transition states. Electrostatic effects will be more pronounced if multiple charges are engineered. It should, therefore, be possible to engineer large effects on both the pH dependence and catalytic rate constants by modifying the surface charge.

This work was supported by the SERC of the UK. We thank Mr J. Brannigan for providing sequencing primers and determining the DNA sequence of the wild-type gene. We also thank Dr Bott for pre-publication data on the distance between His 64 and Asp 99, and Dr J. A. Wells for the sequence of the cloned subtilisin gene prior to publication.

Received 20 August; accepted 3 October 1985.

- Valenzuela, P. & Bender, M. L. *Biochim. biophys. Acta* **250**, 538–548 (1971).
- Perutz, M. *Science* **201**, 1187–1191 (1978).
- Warshel, A., Russell, S. T. & Churg, A. K. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4785–4789 (1984).
- Matthew, J. B., Hanania, G. I. H. & Gurd, F. R. N. *Biochemistry* **21**, 1919–1939 (1979).
- Warshel, A. *Acc. chem. Res.* **14**, 284–290 (1981).
- Markland, F. S. & Smith, E. L. in *The Enzymes* Vol. 3 (ed. Boyer, P. D.) 561–608 (Academic, New York, 1971).
- Kraut, J. in *The Enzymes* Vol. 3 (ed. Boyer, P. D.) 547–560 (Academic, New York, 1971).
- Nedkov, P., Oberthür, W. & Braunitzer, G. *Hoppe-Seyler's Z. physiol. Chem.* **364**, 1537–1540 (1983).
- Smith, E. L., Delange, R. J., Evans, W. H., Landon, M. & Markland, F. S. *J. biol. Chem.* **243**, 2184–2191 (1968).
- Stahl, M. L. & Ferrari, E. *J. Bact.* **158**, 111–118 (1984).
- Kurhara, M., Markland, F. S. & Smit, L. *J. biol. Chem.* **247**, 5619–5631 (1972).
- Gryczan, T. J., Contente, S. & Dubnac, D. *J. Bact.* **134**, 318–329 (1978).
- Jalanko, A., Palva, I. & Söderland, H. *Gene* **14**, 325–328 (1981).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Wells, J. A., Ferrari, E., Henner, D. J., Estell, D. A. & Chen, E. Y. *Nucleic Acids Res.* **11**, 7911–7925 (1983).
- Vasanth, N. et al. *J. Bact.* **159**, 811–819 (1984).
- Kawamura, F. & Doi, R. H. *J. Bact.* **160**, 442–444 (1984).
- Fersht, A. R. *Enzyme Structure and Mechanism* 2nd edn, Ch. 5 (W. H. Freeman, New York, 1985).
- Philipp, M., Tsai, I.-H. & Bender, M. L. *Biochemistry* **18**, 3769–3773 (1979).
- Zoller, M. J. & Smith, M. *Meth. Enzym.* **100**, 468–500 (1983).
- Winter, G., Fersht, A. R., Wilkinson, A. J., Zoller, M. & Smith, M. *Nature* **299**, 756–758 (1982).
- Kramer, B., Kramer, W. & Fritz, H. *J. Cell* **38**, 879–887 (1984).
- Hardy, K. G. in *DNA Cloning, a Practical Approach* (ed. Glover, D. M.) 1–17 (IRL, Oxford, 1985).
- Estell, D. A., Graycar, T. P. & Wells, J. A. *J. biol. Chem.* **260**, 6518–6521 (1985).
- Fersht, A. R. & Renard, M. *Biochemistry* **13**, 1416–1426 (1974).
- Bender, M. L. et al. *J. Am. chem. Soc.* **88**, 5890–5931 (1966).

Loss of genes on the short arm of chromosome 11 in bladder cancer

Eric R. Fearon*, Andrew P. Feinberg*†, Stanley H. Hamilton‡ & Bert Vogelstein*

* The Oncology Center, † Department of Medicine, and ‡ Department of Pathology, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

Recent studies have shown that normal cellular sequences on chromosome 13 are lost during the development of retinoblastomas¹ and that sequences on chromosome 11 are similarly lost during the development of Wilms' kidney tumours²⁻⁵ and embryonal tumours⁶. Cells from these tumours have been found to contain either the paternal or maternal copies of loci on the affected chromosome, but not both. Thus, the somatic loss of heterozygosity for sequences on chromosome 13 or 11 is hypothesized to result in homozygosity for a recessive mutant allele on these chromosomes¹, and in this way the chromosomal loss may contribute to the development of these tumours. We sought to investigate whether similar losses of heterozygosity for chromosome 11 sequences occurred in a common adult tumour. We chose to analyse bladder cancers, since such cancers are common in the adult population and are derived from urogenital tissue, as are Wilms' tumours. We examined constitutional and tumour genotypes at loci on the short arm of chromosome 11 (11p) in 12 patients with transitional cell carcinomas. In five tumours, we observed the somatic loss of genes on 11p resulting in homozygosity or hemizyosity of the non-deleted alleles in the tumour cells. Our results show that the frequency of loss of 11p sequences in bladder cancer approaches that seen in Wilms' tumour (42% compared with 55%)⁷, and suggest that recessive genetic changes involving sequences on 11p may contribute to the development of bladder neoplasms.

The patients studied were 50–85 years old and all had histologically confirmed transitional cell carcinoma (11 of the bladder, 1 of the ureter). All tumour tissues were obtained before any chemotherapy or radiotherapy, and the histological stage and grade of each patient's tumour is indicated in Table 1. DNA of high relative molecular mass was isolated from tumour and normal tissues (either normal bladder mucosa or peripheral leukocytes). To detect the somatic loss of chromosome 11p sequences in the tumours, we exploited DNA polymorphisms on chromosome 11. DNA polymorphisms are normal variations in DNA sequence that occur in the human population and may be used to distinguish the maternal and paternal copies of a gene. The variations can be detected by Southern blotting of

the patient's DNA using a cloned copy of the gene of interest as a probe. Comparison of DNA from normal and tumour tissue can reveal instances where either the paternal or maternal copy of a gene was lost during tumorigenesis.

Two genes on 11p that display DNA polymorphisms are *c-Ha-ras-1* and the insulin gene. *c-Ha-ras-1*, the human homologue of the transforming gene of the Harvey murine sarcoma virus, has been localized⁸ near the distal end of 11p. It displays a length polymorphism due to variations in the number of reiterations of a small tandemly repeated sequence located approximately 1.5 kilobases (kb) 3' to its coding region⁹. Similarly, the insulin gene has been assigned¹⁰ to 11p and it contains an insertion of variable length located approximately 1 kb 5' from the transcriptional start site of the gene¹¹. Figure 1 shows representative Southern blots of DNA from normal and tumour tissues probed with *c-Ha-ras-1* or insulin. Analysis of normal bladder DNA from patient A1 revealed two *c-Ha-ras-1* alleles (denoted 1 and 2 in Fig. 2). No loss of either allele was detected in tumour cells from this patient. Normal tissue from patient A2 was also polymorphic for the *c-Ha-ras-1* gene. However, in this patient's tumour tissue allele 2 was absent (see Fig. 1, lanes A2). Patient D was not heterozygous at the *c-Ha-ras-1* locus, but did display polymorphism at the insulin locus. His normal tissue revealed two insulin alleles (1 and 2 in Fig. 1, lanes D). DNA from his tumour tissue, however, showed significant loss of the hybridization signal from allele 1. Normal tissue from patients M, S2 and T was heterozygous at *c-Ha-ras-1*; in their tumour tissues, these patients had a marked reduction in the intensity of alleles 1, 2 and 1, respectively (see Fig. 1). All 12 patients were heterozygous at either the *c-Ha-ras-1* or the insulin locus. The six patients not shown in Fig. 1 exhibited no loss of 11p alleles in their tumour tissues (see Table 1). There was no apparent correlation between tumour stage or histological grade and the loss of 11p alleles.

To determine whether the losses of 11p alleles in tumour tissue were associated with duplications of the retained alleles, as was previously demonstrated in Wilms' tumours²⁻⁵ and embryonal tumours⁶, we quantified the hybridization signals seen in Fig. 1. At least one Southern blot filter from each of the patients was rehybridized with one or more DNA probes derived from sequences present on chromosomes other than 11. Autoradiographs of these blots were scanned with a densitometer and analysed to determine the ratios of hybridization of restriction fragments obtained from the normal and tumour DNA. In Fig. 2, it can be seen that the hybridization signal of the remaining 11p alleles in patients A2 and D is of approximately double intensity in the tumour cells, and is unchanged in patients M, S2 and T. Thus, the losses of 11p alleles in tumour cells of patients A2 and D were associated

Fig. 1 Representative Southern hybridizations of 11p gene probes to normal tissue (N) and transitional cell carcinoma (C) DNA. Patient designations are shown above the blots, and the numbers shown on the left indicate the polymorphic alleles (numbers correspond to the text and Table 1). Blots from patients A1, A2, M, S2 and T are *MspI* digests of DNA hybridized to the *c-Ha-ras-1* gene probe. DNA from patient D (and sample D') was digested with *HinfI* and hybridized to the insulin gene probe. Note the preservation of heterozygosity at the *c-Ha-ras-1* locus in tumour DNA of patient A1, while tumour DNA from patients A2, D, M, S2 and T reveal loss or marked reduction in the intensity of one allele. In addition, note the increased intensity of the retained allele in patients A2 and D. D' represents DNA extracted from frozen sections of patient D as described in the text.

Methods. DNA of high relative molecular mass was prepared from bladder tumour biopsies and either normal bladder mucosa (patients A1, A2, D, M) or peripheral leukocytes (patients S2, T). The tissues were minced and digested overnight in 1% SDS and proteinase K (0.5 µg ml⁻¹) at 48 °C, and DNA was extracted with phenol and chloroform as described elsewhere¹². DNA concentrations were determined by a diphenylamine assay²¹. Approximately 1 µg of DNA was digested with each restriction enzyme, and the resulting DNA fragments were separated by electrophoresis through agarose gels and transferred to nitrocellulose filters^{22,23}. Gene fragments used as probes were: (1) a 1.0-kb *MspI* *c-Ha-ras-1* genomic fragment containing the polymorphic 3' flanking region²⁴; and (2) a 1.3-kb *PvuII* fragment from the recombinant plasmid hINS 5', containing the polymorphic sequences 5' to the insulin gene²⁵. The gene probe fragments were separated by electrophoresis and oligolabelled directly in agarose to a specific activity of >10⁹ d.p.m. µg⁻¹ (refs 26, 27). Southern hybridization²², post-hybridization washing²⁸ and autoradiography²⁹ were performed as described previously.

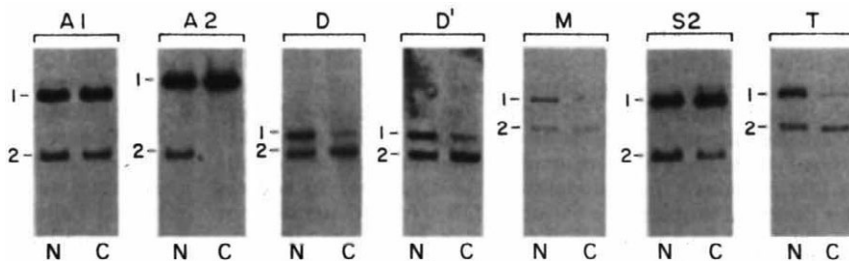


Table 1 Chromosomal alleles in normal tissue (N) or tumour tissue (C) in 12 transitional cell carcinomas

Patient	Stage + grade	1 ATIII	2 CPSI	3 DOSL C9	11 Ha-ras	11 Ins	12 $\alpha 1(II)$	13 p7F12	13 p9D11	13 p9A7	14 pAW101	15 DOSL C3	17 GH	18 DOSL C6	20 DOSL C2
A1 N	T2t3	2/2	3/3	1/1	1/2	1/2	1/1	1/2	1/2	1/1	1/2	1/1	2/2	—	1/2
C		2/2	3/3	1/1	1/2	1/2	1/1	1/2	1/2	1/1	1/2	1/1	2/2	—	1/2/2
A2 N	T3t2	1/2	2/3	1/1	1/2	1/2	1/1	1/2	1/2	1/2	1/2	1/1	1/2	2/2	1/2
C		1/2	2/	1/1	1/1	2/2	1/1	1/2	1/2	1/2	1/2	1/1	1/2	2/2	1/2
B N	T0t1	—	—	1/1	1/2	—	—	2/2	1/1	1/2	—	1/2	1/2	—	1/2
C		—	—	1/1	1/2	—	—	2/2	1/1	1/2	—	1/2	1/2	—	1/2
D N	T3t3	1/2	1/3	—	2/2	1/2	1/1	2/2	1/2	2/2	1/1	1/1	2/2	2/2	1/1
C		1/2	/3	—	2/2	2/2	1/1	2/2	1/2	2/2	1/1	1/1	2/2	2/2	1/1/1
G1 N	T0t2	—	—	1/1	1/2	—	—	1/1	2/2	2/2	—	2/2	2/2	—	1/2
C		—	—	1/1	1/2	—	—	1/1	2/2	2/2	—	2/2	2/2	—	1/2
G2 N	T3t3	1/2	2/3	1/1	1/2	1/1	1/1	2/2	2/2	2/2	1/1	2/2	1/2	2/2	1/1
C		1/2	2/3	1/1	1/2	1/1	1/1	2/2	2/2	2/2	1/1	2/2	1/2	2/2	1/1
M N	T3t3	1/1	3/3	1/1	1/2	—	1/1	—	1/1	1/2	1/2	—	1/2	—	1/1
C		1/1	3/3	1/1	/2	—	1/1	—	1/1/1	1/1/2	1/	—	1/2	—	1/1
P N	T0t2	1/1	2/3	1/1	1/2	1/2	1/1	1/2	1/1	2/2	1/2	1/2	2/2	2/2	1/1
C		1/1	2/3	1/1	1/2	1/2	1/1	1/2	1/1	2/2	1/2	1/2/2	2/2	2/2	1/1
S1 N	T0t1	—	—	—	1/1	1/2	—	1/2	1/2	1/1	—	1/2	1/2	—	1/2
C		—	—	—	1/1	1/2	—	1/2	1/2	1/1	—	1/2	1/2	—	1/2
S2 N	T1t3	1/1	3/3	1/1	1/2	1/2	1/2	1/1	1/2	1/2	1/1	1/2	1/2	2/2	1/2
C		1/1	3/3	1/1	1/	1/	1/2	1/1	1/2	1/2	1/1	1/2	1/2	2/2	1/2
T N	T0t2	2/2	1/3	1/1	1/2	1/2	1/2	1/1	1/2	2/2	1/2	1/2	1/2	1/2	1/2
C		2/2	1/3	1/1	/2	/2	1/2	1/1/1	1/2/2	2/2/2	1/2	1/2	1/2	1/2	1/2
W N	T0t2	—	—	2/2	2/2	1/2	—	1/1	1/1	2/2	—	1/2	2/2	—	1/1
C		—	—	2/2	2/2	1/2	—	1/1	1/1	2/2	—	1/2	2/2	—	1/1

Allelic assignments were determined by Southern hybridizations to 11p gene probes and to probes from chromosomes other than 11p. For each probe, alleles are named with '1' as the larger restriction fragment and '2' or '3' as smaller fragments. '—' Indicates that normal and tumour samples were not analysed with a given probe, due to insufficient amounts of DNA. Chromosomes and the polymorphic probes used to mark these chromosomes are indicated at the top of the table. The clinical stage (T) and histopathological grade (t) of each tumour examined are indicated³². Southern hybridizations as described in the legend to Fig. 1 were performed with the following DNA probes: (1) antithrombin III (ATIII) (chromosome 1q, *Pst*I digest)³³; (2) carbamyl phosphate synthetase (CPSI) (chromosome 2p, *Bgl*I digest)^{13,33}; (3) DOSL C9 (chromosome 3, *Msp*I digest)³⁴; (4) collagen $\alpha 1(II)$ (chromosome 12q, *Hind*III digest)³⁵; (5) p7F12, p9D11, p9A7 (chromosome 13q, *Msp*I digests)³⁰; (6) pAW101 (chromosome 14q, *Eco*RI digest)¹⁴; (7) DOSL C3 (chromosome 15, *Msp*I digest)³⁴; (8) growth hormone (GH) (chromosome 17, *Msp*I digest)³⁶; (9) DOSL C6 (chromosome 18, *Taq*I digest)³⁴; and DOSL C2 (chromosome 20, *Msp*I digest)³⁴. Allele copy number was determined by densitometric analysis.

with duplications of the retained alleles, while a simple loss of 11p genes occurred in tumour cells of patients M, S2 and T, resulting in hemizyosity for 11p sequences.

Densitometric analysis also confirmed the marked reduction in intensity of one allele in each of the patients analysed in Figs 1 and 2. The hybridization signal corresponding to the deleted allele was diminished by 90%, 80%, 90%, 65% and 90% in patients A2, D, M, S2 and T, respectively, when tumour tissue was compared with normal tissue. Complete loss of the signal for the deleted allele was not observed. Two possible explanations for the remaining hybridization signal in the DNA from the tumour specimen were: (1) all tumour cells had lost an allele, but contaminating non-tumour cells were present within the tumour; or (2) the majority, but not all, of the tumour cells had lost an allele. In support of the first possibility, histological sections of tumour specimens from these patients revealed numbers of stromal cells roughly consistent with the signal remaining in the band corresponding to the deleted allele. However, possible cellular heterogeneity within the tumours created difficulty in ascertaining that the piece of tumour tissue used to prepare DNA had the same proportion of non-tumour and tumour cells as that examined for histopathological diagnosis. To address this potential problem, we extracted DNA from tissue sections in which the relative DNA contribution of tumour and non-tumour cells could be precisely determined. A piece of the tumour from patient D was frozen in OCT compound and 100 serial 6- μ m sections were cut with a cryostat. Sections 1, 51 and 99 were stained with haematoxylin and eosin for histological analysis and the remainder of the sections were used to prepare DNA using a technique described recently¹². The Southern blot obtained with the DNA from these sections is shown in Fig. 1, lanes D'. The result is quite similar to that obtained with the DNA isolated from a large piece of this tumour (Fig. 1, lanes D). Densitometry of the autoradiograph showed that the signal corresponding to the deleted allele was diminished by 83% when the DNA from the tumour sections was compared with

normal tissue from the same patient. Histological analysis revealed that 67% $\pm$ 15% of the cells within these sections (3,800 cells in 26 fields examined) were tumour cells. Section 100 was stained with Feulgen, and microdensitometry was performed using a Leitz DAD video analysis system. This analysis revealed that the normal cells within the tumour had a relative DNA content of 1.0 ± 0.15 while the tumour cells had a relative DNA content of 1.29 ± 0.33 . Hence, we expected that 86% ($1.29 \times 67\%$) of the DNA extracted from these sections would be derived from tumour cells, and 14% derived from stromal cells within the tumour tissue. Comparison of the results of histological analysis and Southern blot analysis suggests that the residual hybridization of allele 1 in the tumour specimen from patient D was largely derived from contaminating stromal cells.

To obtain evidence that the changes we observed were specific to 11p and not simply widespread random alterations, we also examined DNA markers on chromosomes 1-3, 12-15, 17, 18 and 20. These data are summarized in Table 1. Interestingly, some changes were noted when genes from normal and tumour DNAs were compared on these chromosomes, but the incidence of these changes was less frequent than those seen on 11p. For example, we detected the loss of one carbamyl phosphate synthetase I allele (chromosome 2p)¹³ in each of the tumours from patients A2 and D (Table 1), and the loss of one pAW101 allele (chromosome 14q)¹⁴ in patient M. In addition, while reduction to homozygosity for chromosome 13 markers occurs in 60% of retinoblastomas¹, we did not observe the loss of any chromosome 13 markers in bladder tumours. Tumour tissues from patients M and T did exhibit a duplication of sequences on one chromosome 13 homologue, resulting in triploidy for chromosome 13 sequences (Table 1). Similarly, comparison of DNA obtained from tumour tissues of patients A1, D and P with DNA from their normal tissues showed hybridization intensities consistent with triploidy for chromosome 20 or chromosome 15 sequences in the tumour tissues. No other qualitative or quantitative chromosomal changes were detected in the other chromosomes

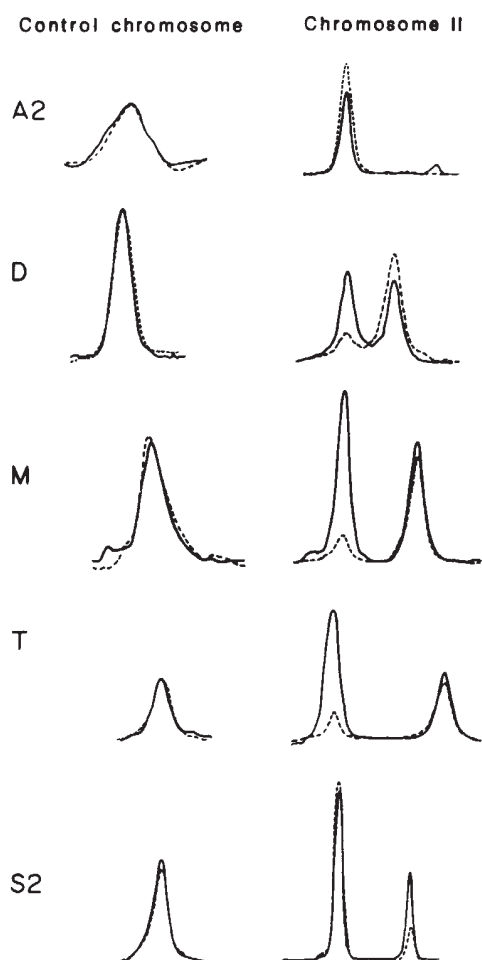


Fig. 2 Densitometric analyses of Southern hybridizations with 11p gene probes and of rehybridizations with non-chromosome 11 genes (control chromosome). Solid lines, normal lanes; broken lines, tumour lanes. Patient designations are indicated to the left. **Methods.** Southern blots (from Fig. 1) were freed of probe in alkali and rehybridized with control genes: (1) p9A7, a recombinant plasmid containing 0.95 kb of sequences derived from chromosome 13q³⁰ (patients A2, D and M; allele 2 was used for comparisons between the normal and tumour tissues in all three patients); or (2) pEP2, a recombinant plasmid containing coding sequences from the ATIII gene on chromosome 1q³¹ (patients S2 and T). Grain density was quantified using a Clifford Densitomp model 445 densitometer. After normalizing the control scans for DNA loading, the 11p gene blots were re-scanned to produce the tracings shown.

examined (see Table 1). Hence, while loss of 11p sequences is certainly not the only genetic event that occurs in bladder cancer, our data would argue that such loss is relatively frequent in this cancer, and may be one of the important steps involved in development of this tumour.

We were unable to perform karyotypic analyses of the tumours studied. However, previous karyotypic analyses^{15,16} of transitional cell bladder carcinomas have revealed alterations of chromosome 1, 5, 8, 9, 11 and 13. Significantly, chromosomal losses involving 11p were seen in a subfraction of tumour cells from 3 of 9 patients in one study¹⁵ and at least 3 of 10 patients in another¹⁶. We observed a higher incidence of 11p losses than that revealed by karyotypic analysis. In addition, our results indicated that loss of 11p sequences occurred in the majority of the tumour cells from 5 of 12 patients. Thus the situation with bladder tumours is similar to that seen in Wilms' tumours, wherein analysis of DNA polymorphisms reveals a higher incidence of loss of 11p sequences than usually seen karyotypi-

cally¹⁷. These differences between the two analyses may be attributed to: (1) the difficulties inherent in cytogenetic studies of complex karyotypes from solid tumour specimens; (2) karyotypic analysis may be biased towards cells with the highest mitotic indices, while DNA polymorphism analysis represents an averaging of all cells within the tumour specimen; and (3) abnormal chromosome segregation which produces loss and reduplication events without gross loss of chromosomal material (such as that likely to have occurred in patients A2 and D) will not be detected karyotypically.

The chromosome 11 losses we observed were probably important for some stage of tumorigenesis in the five cases in which they occurred. This conclusion is strongly supported by our observation that the chromosomal loss, when observed, appeared to be present in nearly all cells of the tumour specimens (see Fig. 2). Hence, the losses did not simply occur in a small subclone of the tumour cell population; rather, the clone that lost 11p sequences apparently had a selective advantage sufficient to allow it to outgrow virtually all other cells.

There are at least two possible explanations for the growth advantage associated with loss of 11p sequences. First, it is possible that a normal cellular constraint of growth is present on chromosome 11p, and that the loss of this suppressor locus leads to activation of a normally repressed class of genes¹⁸. This mechanism does not seem to involve a simple dosage phenomenon. Bladder tumours from patients A2 and D and all of the Wilms' tumours in previous studies had a normal dosage of chromosome 11p arms, even though this diploid dosage was due to the loss of one chromosome 11p and the duplication of its homologue. Hence, for this explanation to be correct, one must postulate that two events had occurred in these cases: an inactivation of one chromosome 11p suppressor locus (by mutation, for example) followed by a loss of the chromosome 11p containing the normal suppressor locus, sometimes accompanied by the duplication of the abnormal 11p^{1,19}.

An alternative explanation for our results is that the events involving loss of 11p are simply correlated with another event that results in loss of growth control. In this model, the loss of 11p would itself have no intrinsic effect on the development of the tumour. For example, if a cell was prone to aneuploidy and chromosome 11 losses were not lethal to the cell, then some cells with a loss of chromosome 11 might also contain other alterations that could impart a growth advantage. This cell could then proliferate to produce a tumour with a clonal loss or loss and reduplication of 11p sequences. The fact that loss of 11p is only noted in 40–60% of bladder cancers and Wilms' tumours is consistent with this hypothesis.

In this context, however, examination of other tumours by molecular hybridization techniques does not reveal frequent loss of 11p. Of eight retinoblastomas studied for loss of 11p sequences, none showed loss of 11p^{1,20}. In addition, of 15 colon tumours examined, loss of 11p sequences was seen in only one tumour (E.R.F. and B.V., unpublished observations). Thus, loss of chromosome 11 sequences occurs frequently, and relatively specifically, in transitional cell carcinomas, Wilms' tumours, and other embryonal tumours. Its role in tumorigenesis will remain speculative, however, until the advent of studies which reveal the molecular effect of such loss.

We thank Dr Gunter F. Bahr (Armed Forces Institute of Pathology) for microdensitometry analysis, Dr Michael Droller for providing the tissues used in this study and Dr Susan Goelz for critical discussions. We are also grateful to the following investigators for DNA probes: Dr R. Weinberg for c-Ha-ras-1; Dr P. Rotwein for insulin; Dr P. O'Connell and Dr R. White for chromosome 13 sequences and DOSLC 2, 3, 6 and 9; Dr E. Prochownik for antithrombin III; Dr W. O'Brien for carbamyl phosphate synthetase I; Dr C. Francomano and A. Nunez for collagen 1(II); Dr S. Orkin for pAW101; and Dr P. Seeburg for growth hormone. We also thank Sandy Smith for typing the manuscript. This work was supported by the Clayton Fund, the Hartford Foundation (A.P.F.), training grant GM 07309 (E.R.F.), and NIH grant CA35494.

Received 14 March; accepted 25 September 1985.

1. Cavenee, W. K. *et al.* *Nature* **305**, 779-784 (1983).
2. Koufos, A. *et al.* *Nature* **309**, 170-172 (1984).
3. Orkin, S. H., Goldman, D. S. & Sallan, S. E. *Nature* **309**, 172-174 (1984).
4. Reeve, A. E. *et al.* *Nature* **309**, 174-176 (1984).
5. Fearon, E. R., Vogelstein, B. & Feinberg, A. P. *Nature* **309**, 176-178 (1984).
6. Koufos, A. *et al.* *Nature* **316**, 330-334 (1985).
7. Solomon, E. *Nature* **309**, 111-112 (1984).
8. Fisher, J. H. *et al.* *Somat. Cell Molec. Genet.* **10**, 455-464 (1984).
9. Capon, D. J., Charen, E. Y., Levinson, A. D., Seeburg, P. H. & Goeddel, D. V. *Nature* **302**, 33-37 (1983).
10. Harper, M. E., Ullrich, A. & Saunders, G. A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4458-4460 (1981).
11. Bell, G. I., Selby, M. J. & Rutter, W. J. *Nature* **295**, 31-35 (1982).
12. Goeltz, S. E., Hamilton, S. R. & Vogelstein, B. *Biochem. biophys. Res. Commun.* **130**, 118-126 (1985).
13. Adcock, M. W. & O'Brien, W. P. *J. biol. Chem.* **259**, 13471-13476 (1984).
14. Wyman, A. R. & White, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6754-6758 (1980).
15. Gibos, Z., Prout, G. R., Connolly, J. G., Pontes, J. E. & Sandberg, A. A. *Cancer Res.* **44**, 1257-1264 (1984).
16. Atkin, N. B. & Baker, M. C. *Cancer Genet. Cytogenet.* **15**, 253-268 (1985).
17. Kondo, K., Chilcote, R. R., Mauver, H. S. & Rowley, J. D. *Cancer Res.* **44**, 5376-5381 (1984).
18. Comings, D. E. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3324-3328 (1973).
19. Knudson, A. G. & Strong, L. C. *J. natn. Cancer Inst.* **48**, 313-324 (1972).
20. Dryja, T. P. *et al.* *New Engl. J. Med.* **310**, 550-553 (1984).
21. Burton, K. *Biochem. J.* **62**, 315-323 (1956).
22. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
23. Wahl, G. M., Stern, M. & Stark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3683-3687 (1979).
24. Shih, C. & Weinberg, R. A. *Cell* **29**, 161-167 (1982).
25. Rotwein, P. S. *et al.* *New Engl. J. Med.* **308**, 65-71 (1983).
26. Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6-13 (1983).
27. Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **137**, 266-267 (1984).
28. Peden, K., Mounts, P. & Hayward, G. S. *Cell* **31**, 71-80 (1982).
29. Swanson, R. & Shank, P. *Analyt. Biochem.* **86**, 184-192 (1978).
30. Cavenee, W., Leach, R., Mohandas, T., Pearson, P. & White, R. *Am. J. hum. Genet.* **36**, 10-24 (1984).
31. Prochownik, E. V. *et al.* *New Engl. J. Med.* **308**, 1549-1552 (1983).
32. Robbins, S. L., Cotran, R. S. & Kumar, V. *Pathologic Basis of Disease* 3rd ed 1070-1076 (Saunders, Philadelphia, 1984).
33. Fearon, E. R. *et al.* *Hum. Genet.* **70**, 207-210 (1985).
34. Skolnick, M. A., Willard, H. F. & Menlove, L. A. *Cytogenet. Cell Genet.* **37**, 210-273 (1984).
35. Francomano, C. A., Nunez, A., Yamada, Y. & Phillips, J. A. *Am. J. hum. Genet.* **37S**, 404 (1984).
36. Phillips, J. A. *et al.* *J. clin. invest.* **70**, 489-495 (1982).

A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7

Robert G. Knowlton*, Odile Cohen-Haguenauer†, Nguyen Van Cong†, Jean Frézal†, Valerie A. Brown*, David Barker*, Jeffrey C. Braman*, James W. Schumm*, Lap-Chee Tsui‡, Manuel Buchwald‡ & Helen Donis-Keller*

* Department of Human Genetics, Collaborative Research, Inc., 128 Spring St., Lexington, Massachusetts 02173, USA

† Unité de Recherches de Génétique Médicale (INSERM U12), Hôpital des Enfants Malades, 149 rue de Sèvres, F-75743 Paris, Cedex 15, France

‡ Department of Genetics, The Hospital for Sick Children, and Department of Medical Genetics and Medical Biophysics, University of Toronto, Toronto, Ontario, Canada M5S 1A1

Although cystic fibrosis (CF) is among the most common inherited diseases in Caucasian populations¹, the basic biochemical defect is not yet known. CF is inherited as an autosomal recessive trait apparently due to mutations in a single gene²⁻⁴, whence the efforts made to identify the genetic locus responsible by linkage studies. Two markers have recently been identified that are genetically linked to CF: one is a genetic variation in serum level of activity of the enzyme paraoxonase⁵, and the other is a restriction fragment length polymorphism (RFLP) identified with a randomly isolated DNA probe⁶. We report here that the genetic locus DOCRI-917 defined by the cloned DNA probe is located on chromosome 7.

The polymorphic locus DOCRI-917 is detected with the probe LAM4-917, a phage clone from the genomic library of Lawn *et al.*⁶ which is one of a set of probes identified by a random screening procedure for RFLPs⁷. It is a 17-kilobase (kb) single-copy human genomic DNA sequence polymorphic both at a *Hind*III site and a *Hinc*II site in Southern blot hybridizations. The probe is frequently informative in inheritance studies, having a polymorphism information content⁸ of 0.57.

The structure and inheritance of the locus are described by Tsui *et al.*⁴.

The marker locus DOCRI-917 was identified by DNA hybridization in rodent-human hybrid cell lines containing different human chromosome complements⁷. Hybridization of the probe to DNA from 17 hamster-human hybrids, 8 mouse-human hybrids and the parental human and rodent cell lines is shown in Fig. 1. As expected from the restriction map of the DOCRI-917 locus⁴, four fragments (8.9, 5.8, 5.0 and 3.6 kb) are detected in all lanes with complementary genomic sequences. (Restriction fragment length polymorphism at the DOCRI-917 locus is not observed with *Eco*RI.) The probe does not cross-hybridize with mouse or hamster sequences (lanes CH and M).

Comparison of the LAM4-917 hybridization results with the chromosome content of the 25 cell lines indicates that the DOCRI-917 locus is situated on chromosome 7. Probe hybridization conforms to the expected pattern of a chromosome 7 locus in every line, and shows multiple discordances with all other chromosomes (Table 1). In one cell line (CH13) that scored positive with the LAM4-917 probe, chromosome 7 was present at an earlier passage but was no longer detectable by karyotype analysis or by assay of the β -glucuronidase isozyme marker GUSB¹⁰. At least some chromosome 7 sequences were present, however, as shown by hybridization of CH13 DNA with a complementary DNA clone of the T-cell receptor β -chain gene¹¹, located on chromosome 7 (q3) (refs 12-15). As shown in Table 1, the hybridization results of LAM4-917 to all 25 lines were identical to those obtained with the T-cell receptor β -chain probe.

All chromosomes except chromosome 7 are excluded because they are present in one or more of the cell line DNAs not hybridizing to LAM4-917. It is particularly important to note that chromosome 2, even though the number of discordances is low, is excluded as the location of DOCRI-917 by the absence of hybridization of the probe in the line CH8. Chromosome 2 was detected in karyotype analysis of CH8 metaphase cells, the cells were positive for the chromosome 2 isoenzymes MDH1 and IDH1¹⁶, and hybridization to a DNA probe from chromosome 2 (D2S1¹⁷) is clearly detected in the same CH8 DNA sample. Exclusion of DOCRI-917 from the X and Y chromosomes was also confirmed by hybridization to a set of cell lines with sex chromosome aneuploidies (results not shown).

Because of its established genetic linkage to DOCRI-917, we conclude that the cystic fibrosis locus is also situated on chromosome 7. Determination of the location of the cystic fibrosis gene is an important step in identifying the primary genetic defect responsible for the disease. Any hypothesis that a particular genetic function is the primary lesion must satisfy the criterion of mapping to the same site as the cystic fibrosis trait. With the discovery of linkage between CF and the genetic markers PON⁵ and DOCRI-917⁶, the location of the disease gene is restricted to the 1% of the genome surrounding these markers. By chromosomal localization of DOCRI-917, that region is now mapped to chromosome 7. More specific localization of DOCRI-917 on chromosome 7 will be possible by testing other hybrid cell lines containing partial deletions of chromosome 7, *in situ* hybridization to metaphase chromosomes, and linkage mapping with other chromosome 7 markers. The observation that the DOCRI-917 locus is present in cell line CH13, which appears to retain only a part of chromosome 7 lacking GUSB (cen→q22) (ref. 18) but retaining the T-cell receptor β -chain gene TCRB on 7q3 is consistent with a location on 7q, but other explanations of this result are possible. For example, more than one segment of chromosome 7 might have segregated with this line, or the DNA hybridization may simply be more sensitive than the other methods used to detect chromosome 7.

Population studies have supported the hypothesis that mutations in a single autosomal gene are responsible for CF^{2,3}. In addition, the observation of only 15% recombination between CF and the marker DOCRI-917 in 39 families indicates that the disease is attributable to mutations at the site we have identified on chromosome 7 in the vast majority of these families⁴. At

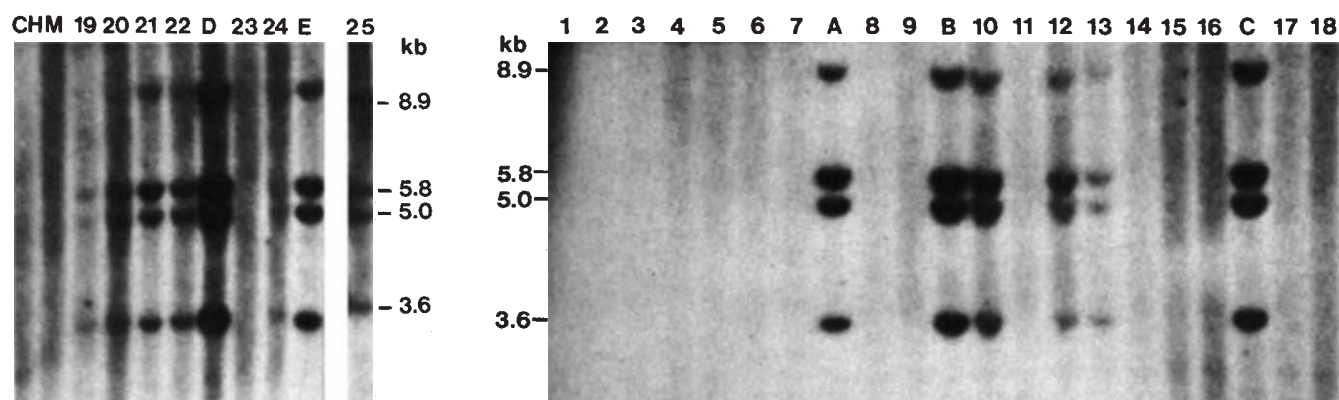


Fig. 1 Autoradiogram of hybridization of LAM4-917 to DNA of 25 somatic cell hybrids. Genomic DNA prepared from each hybrid cell line was digested with *EcoRI*, separated and transferred to DBM-filters¹⁹. DNA of the phage clone LAM4-917 was radioactively labelled with ³²P by nick-translation to a specific activity of 2×10^8 d.p.m. μg^{-1} . Hybridization was carried out for 20 h at 42 °C in 50% formamide, 0.6 M NaCl, 0.05 M Tris-HCl pH 7.6, 0.1% sodium pyrophosphate, 0.1% SDS, 0.04% Ficoll, 0.04% polyvinylpyrrolidone, 0.04% bovine serum albumin, 5% dextran sulphate, and 250 $\mu\text{g ml}^{-1}$ sonicated, denatured salmon sperm DNA. The filters were washed in $2 \times \text{SSC}$ at 20 °C for 15 min and twice at 65 °C for 15 min, in $0.2 \times \text{SSC}$, 0.2% SDS at 65 °C for 40 min, and finally in $0.1 \times \text{SSC}$, 0.1% SDS at 65 °C for 40 min. The filters were exposed to Kodak XAR-5 film with intensifying screens (Dupont Cronex Lightning Plus) at -70 °C. The lane numbers (1-25) correspond to the hybrid cell lines listed in Table 1. Lanes 1-17 are hamster-human hybrids, and lanes 18-25 are mouse-human hybrids. A-E are the human parental cell lines, CH is the chinese hamster cell line V79.4, and M is the mouse line C11DA.

Table 1 Chromosome localization of DOCRI-917

Cell line	Human chromosomes																						Translocation chromosomes	Hybridization				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22		X	Y	TCRB	DOCRI-917	
1. CH1	-	-	-	+	-	+	-	+	/	+	-	+	+	/	-	-	-	/	-	/	+	-	-	-	-	Xp/2q	-	-
2. CH2	-	-	+	+	-	-	-	+	-	-	-	-	-	+	-	-	/	-	-	/	+	-	-	-	+	Xp/2q	-	-
3. CH3	+	-	/	-	-	+	-	+	+	/	+	+	+	-	+	+	-	+	-	-	+	-	-	-	-	Xp/2q	-	-
4. CH4	+	-	+	+	+	/	-	+	+	-	+	+	+	+	/	-	-	+	+	+	-	-	-	-	-	Xp/2q	-	-
5. CH5	-	-	-	-	-	+	/	/	-	+	+	+	-	+	-	/	-	-	-	+	+	+	-	-	-	Xp/2q	-	-
6. CH6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	/	-	-	-	+	-	-	-	-	Xp/2q	-	-
7. CH7	-	-	+	+	+	-	-	+	+	-	+	+	-	+	+	-	+	+	+	+	+	-	-	-	-	-	-	-
8. CH8	+	+	+	/	+	+	-	+	+	-	-	+	-	-	-	-	-	+	+	-	-	+	-	-	-	Xp/5q, 5p/Xq	-	-
9. CH9	-	-	-	-	-	-	-	+	-	-	-	-	-	+	+	+	-	+	+	-	+	-	+	-	-	-	-	-
10. CH10	+	-	-	+	+	+	+	-	+	-	+	-	-	-	-	-	+	/	+	+	+	+	+	+	-	+	+	
11. CH11	-	-	+	-	+	+	/	-	-	/	+	-	+	-	+	-	-	/	-	/	+	+	+	-	-	Xp/2q	-	-
12. CH12	-	-	-	+	+	+	+	/	-	-	+	+	+	+	+	+	-	+	+	-	+	/	+	-	-	Xp/2q	+	+
13. CH13	-	-	-	+	-	-	/*	/	-	-	+	+	-	-	-	-	-	+	+	-	-	-	+	-	-	Xp/2q	+	+
14. CH14	+	-	+	+	-	+	-	-	+	/	+	+	+	+	+	+	-	+	+	+	+	+	+	+	-	Xp/2q	-	-
15. CH15	-	-	-	+	+	+	-	/	-	+	+	-	+	+	-	/	-	-	+	-	+	-	-	/	-	Xp/2q	-	-
16. CH16	-	-	+	/	-	+	-	/	/	/	-	-	/	+	+	+	-	/	+	-	-	-	-	-	-	Xp/2q	-	-
17. CH17	-	-	-	-	-	+	-	+	-	-	+	+	-	-	-	-	-	-	-	+	/	/	+	-	-	-	-	-
18. M1	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	-	-	-	/	+	+	-	+	/	-	-	-	-
19. M2	+	+	-	+	+	-	+	+	-	/	+	+	-	-	+	+	+	+	+	+	+	+	+	/	-	Xp/5q	+	+
20. M3	-	+	+	/	+	-	+	+	-	-	+	+	/	-	+	+	+	+	-	+	+	+	+	-	-	5p/Xq	+	+
21. M4	/	+	+	-	+	+	+	+	-	+	+	+	+	-	+	+	+	+	+	+	+	+	+	-	-	+	+	
22. M5	-	+	+	/	+	+	+	-	-	+	-	-	-	-	-	-	-	+	/	+	+	+	+	/	-	+	+	
23. M6	-	-	+	/	-	+	-	+	-	-	-	+	+	+	-	/	+	-	+	+	+	+	-	/	-	-	+	-
24. M7	/	/	+	+	+	/	+	+	-	-	+	-	+	/	-	-	+	-	+	/	-	-	/	-	-	+	+	
25. M8	-	+	+	-	-	+	+	-	-	-	+	+	-	-	-	-	+	-	-	-	/	-	+	-	-	Xp/5q, 5p/Xq	+	+
Discordant																												
DOCRI-917 (%)	39	17	50	40	28	57	0	60	57	55	36	48	50	68	54	43	21	29	48	45	59	29	32			0		

Chromosomal content of hybrid cell lines was determined by cytogenetic and isoenzyme analysis. The panel of hybrid cell lines was generated and characterized by Nguyen Van Cong, Dominique Weil and Catherine Finaz, Unité de Recherches de Génétique Médicale, INSERM U12, Paris, France, and is described elsewhere²⁰. Symbols: +, chromosome detected in more than 30% of cells; -, chromosome not detected; /, chromosome detected in less than 30% of cells, not scored for mapping. Three of the human parental cell lines contained reciprocal chromosomal translocations^{20,21} retained by some hybrids: Xq/2q (Xqter → p22:2q32 → 2qter), Xp/5q (Xppter → q21:5q11 → qter), and 5p/Xq (5pter → q11:Xq21 → qter). Hybridization results for DOCRI-917 are from the autoradiogram in Fig. 1. The TCRB gene is detected by hybridization with the T-cell receptor β -chain probe¹¹ with the same cellular DNA samples. * Possible chromosome rearrangement, see text.

present, we cannot exclude the possibility that mutations in genes not linked to DOCRI-917 and PON are responsible for CF in a small percentage of families.

The localization of DOCRI-917 to chromosome 7 greatly facilitates the identification of other RFLP markers linked to CF. Polymorphic probes already assigned to chromosome 7 can be tested for linkage to CF, and collection of additional RFLP markers for this purpose can be accelerated by drawing probes from chromosome 7-specific libraries. Such RFLP markers will

contribute to increasing the resolution of the linkage map of the CF region, and ultimately to the identification of the CF gene itself.

We thank Marie-France de Tand, Chantal Foubert and Marie-Sylvie Gross for their technical assistance.

Note added in proof: Our recent linkage analysis shows that both CF and DOCRI-917 are closely linked to the pro α 2(1) collagen gene (7q21-q22) (ref. 22 N. Buchwald et al., in preparation).

Received 11 November; accepted 12 November 1985.

1. Talamo, R. C., Rosenstein, B. J. & Berninger, R. W. in *The Metabolic Basis of Inherited Disease* 5th edn (eds Stanbury, J. B. et al.) 1889–1917 (McGraw-Hill, New York, 1983).
2. Danks, D. M., Phelan, P. D. & Chapman, C. *Am. J. hum. Genet.* **36**, 1389–1402 (1984).
3. Romeo, G. et al. *Am. J. hum. Genet.* **37**, 338–349 (1985).
4. Tsui, L.-C. et al. *Science* (in the press).
5. Schmiegelow, K. et al. *Lancet* (in the press).
6. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. *Cell* **15**, 1157–1174 (1978).
7. Schumm, J. W. et al. *8th Human Gene Mapping Workshop*, Helsinki (1985); *Cytogenet. Cell Genet.* (in the press).
8. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. *Am. J. hum. Genet.* **32**, 314–331 (1980).
9. Nguyen Van Cong, Weil, D. & Finaz, C. *Annls Génét.* (in the press).
10. Grzeschik, K. H. *Somatic Cell Genet.* **2**, 401–410 (1976).
11. Jones, N. et al. *Science* **227**, 311–314 (1985).
12. Barker, P. E., Ruddle, F. H., Royer, H.-D., Acuto, O. & Reinherz, E. L. *Science* **226**, 348–349 (1984).
13. Collins, M. K. L. et al. *EMBO J.* **3**, 2347–2349 (1984).
14. Isobe, M., Erikson, J., Emanuel, B. S., Nowell, P. C. & Croce, C. *Science* **228**, 580–582 (1985).
15. Morton, C. C., Duby, A. D., Eddy, R. L., Shows, T. B. & Seidman, J. G. *Science* **228**, 582–584 (1985).
16. Shows, T. B. *Biochem. Genet.* **7**, 193–204 (1972).
17. Gedde-Dahl, T. et al. *8th Human Gene Mapping Workshop*, Helsinki (1985); *Cytogenet. Cell Genet.* (in the press).
18. Meera Khan, P. & Robson, E. B. *Human Gene Mapping* **5**, 39–46 (1979).
19. Wahl, G. M., Stern, M. & Stark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3683–3687 (1979).
20. Gross, M.-S. et al. *Annls Génét.* **25**, 87–91 (1982).
21. Weil, D. et al. *Annls Génét.* **24**, 89–92 (1981).
22. Retief, E., Parker, M. I. & Retief, A. E. *Hum. Genet.* **69**, 304–308 (1985).

A closely linked genetic marker for cystic fibrosis

Ray White*, Scott Woodward*, Mark Leppert*,
Peter O'Connell*, Mark Hoff*, John Herbst*,
Jean-Marc Lalouel*, Michael Dean†
& George Vande Woude‡

* Howard Hughes Medical Institute and the Cystic Fibrosis Research Center, Department of Human Genetics and † Department of Pediatrics, University of Utah College of Medicine, Salt Lake City, Utah 84132, USA

‡ Litton Bionetics, Inc., Basic Research Program, NCI-Frederick Cancer Research Facility, Frederick, Maryland 21701, USA

Cystic fibrosis is a recessive genetic disorder, characterized clinically by chronic obstructive lung disease, pancreatic insufficiency and elevated sweat electrolytes; affected individuals rarely live past their early twenties. Cystic fibrosis is also one of the most common genetic diseases in the northern European population. The frequency of carriers of mutant alleles in some populations is estimated to be as high as 1 in 20, carrying a concomitant burden of about one affected child in 1,500 births. Because little is known of the essential biochemical defect caused by the mutant gene, a genetic linkage approach based on arbitrary genetic markers and family studies is indicated to determine the chromosomal location of the cystic fibrosis (CF) gene. We have now obtained evidence for tight linkage between the CF locus and a DNA sequence polymorphism at the *met* oncogene locus. This evidence, combined with the physical localization data for the *met* locus presented in the accompanying paper¹, places the CF locus in the middle third of the long arm of chromosome 7, probably between bands q21 and q31.

Appropriate families are required for genetic linkage analyses with recessive disease loci. For this study, we sampled and tested 13 families with multiple affected offspring. Siblings were characterized at cystic fibrosis clinical centres as affected or unaffected based on quantitative sweat chloride determination as well as other clinical features. Several of the families were from Utah, but most came from different regions in the United States.

We estimate that more than 100 loci have been tested for linkage in CF, including 21 probes from our own laboratory, with negative results. However, Eiberg *et al.* have reported evidence of linkage between the CF locus and the gene for the enzyme paroxonase², and recently Tsui and others have obtained evidence for genetic linkage with a DNA marker (L.-C. Tsui, personal communication). Both of these linkages are rather

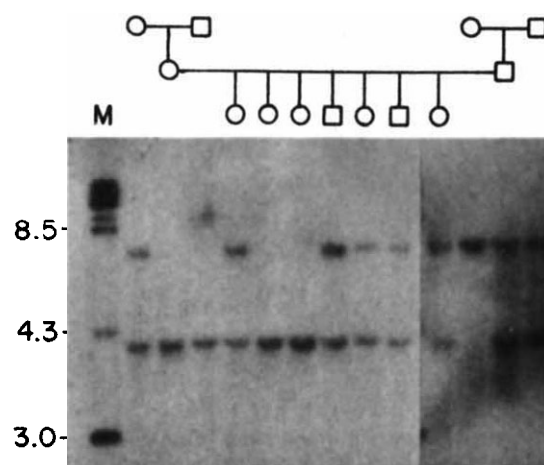


Fig. 1 Segregation of the *TaqI* polymorphism in a complete three-generation family. DNAs from the indicated individuals were digested with the restriction enzyme *TaqI*, electrophoresed in agarose and transferred by the method of Southern to a nylon filter (Micron Separations, Inc.). The filters were hybridized⁹ with the clone *metH* (refs 1, 3). The polymorphisms were found by examination of a panel of DNAs from six unrelated individuals digested with the restriction enzymes *MspI*, *TaqI*, *EcoRI*, *HindIII*, *PstI* or *BclI*. Only *MspI* and *TaqI* revealed polymorphism. Individuals digested with *TaqI* yielded fragment lengths of either 7.5 kb (allele 1) alone or 4.0 kb (allele 2) alone, or both 7.5 and 4.0 kb. Digestion with *MspI* revealed individuals with fragment lengths of 2.3 and 1.8 kb, or only one of the two. In six informative families with large sibships, we found no exceptions to mendelian inheritance. A survey of 60 unrelated individuals gave a frequency of 0.56 for the 7.5-kb *TaqI* allele and the 2.3-kb *MspI* fragment, and 0.44 for each of the two smaller fragments. In all cases where phase could be distinguished, the large fragments were found together as a single haplotype, indicating a high degree of linkage disequilibrium in this marker system.

loose (10% and 15% recombination, respectively) but they do begin to pave the way for a more precise localization. Klinger has also reported suggestive, but not definitive, evidence for linkage of cystic fibrosis to a genetic marker located on the short arm of chromosome 21 in two large Amish pedigrees. However, the linkage was not evident when a collection of outbred nuclear families was examined (Klinger, personal communication). To develop a genetic marker at the *met* locus, we cloned fragment H (refs 1, 3), a 1.6-kilobase (kb) *SalI/EcoRI* fragment, into pBR322. Figure 1 shows the inheritance of a polymorphic *TaqI* restriction fragment revealed by this probe in a complete three-generation family. An *MspI* fragment also showed polymorphism with this probe. Of 60 unaffected unrelated individuals examined, 36 were heterozygous. However, we have seen only two haplotypes thus far.

At least one parent was heterozygous at the *met* locus in 12 of the 13 CF families investigated. Linkage tests between the CF locus and the *met* locus were performed with the computer program LINKAGE (ref. 4). Our results are reported in Fig. 2 and Table 1. As shown in Fig. 2, the LOD score reaches a maximum of 8.65 at a recombination value of 0, indicating no evidence for recombination. A LOD score of 8.65 corresponds to odds of $4 \times 10^8:1$, favouring the hypothesis of complete linkage over that of independent segregation. We also verified that in each family the LOD score is maximal under the hypothesis of complete linkage. As in all estimation situations, our sample estimate of the recombination rate between these two loci may depart from the true, unknown value because of sampling error. As indicated in Fig. 2, however, the true recombination frequency is unlikely to be $>5\%$. Our data therefore indicate tight linkage between cystic fibrosis and the *met* gene in these families.

Table 1 shows the composition, the genotypes at the *met* locus and the LOD score for each family. Under the hypothesis of

Table 1 Summary of family data

Family	Parents	Offspring	LOD score
1409	12 × 11	3A (11)	0.60
1414	12 × 22	3A (12)	0.60
1415	12 × 12	3A (11) 1N (12)	1.33
1422	12 × 11	3A (12)	0.60
1425	12 × 11	5A (12) 2N (12, 11)	1.15
1426	22 × 12	3A (12)	0.60
1427	12 × 12	3A (11) 4N (12)	1.70
1438	11 × 12	2A (11)	0.30
1442	12 × 11	3A (11)	0.60
1446	11 × 12	3A (11)	0.60
774	11 × 12	2A (11)	0.30
1378	22 × 12	3A (22) 2N (22)	0.25
1436	11 × 22	3A (12)	0.00

Genotypes at the marker locus reported for each family member. The LOD score for each informative family at a recombination value of 0 is also reported, allowing a closer inspection of the contribution of each family to the linkage evidence. Assuming that all parents are heterozygous at the CF locus, all but two (families 1415, 1427) are single backcrosses with respect to the test locus. In such families, each affected offspring contributes terms equal to θ and $1 - \theta$ under each parental phase respectively. Under complete linkage, phase information can be derived from a single affected individual. It follows that each additional individual will contribute 2 units to the odds ratio, or equivalently 0.301 to the LOD score, while the penalty of having to infer phase amounts to 0.301 LOD unit in each family. In two instances of such crosses, unaffected siblings were available for testing. In family 1425, five affected siblings have genotype 12, whereas two unaffected siblings have genotypes 12 and 11, respectively. Without the unaffected siblings, the LOD score would have been 1.20 instead of 1.15. The two unaffected siblings, while compatible with the hypothesis of no recombination, have discordant genotypes at the test locus and contribute terms $(2 - \theta)(1 + \theta)$ under each parental phase, accounting for -0.05 in LOD units. In family 1378, the contribution of three affected offspring is offset by two unaffected siblings with the same genotype at the *met* locus, each unaffected sibling accounting for terms $(2 - \theta)$ and $(1 + \theta)$ under each phase. Family 1427 is a double intercross and contributes more support to the hypothesis of complete linkage than any other family. This results from the increased contribution of affected individuals of genotype 11 in this intercross as opposed to the backcrosses, as well as from the information provided by the four unaffected siblings of genotype 12. Once an affected individual has yielded phase, each additional affected sibling of genotype 11 contributes $(1 - \theta)^2$, or 0.602 LOD units, while each unaffected sibling of genotype 12 contributes $2(1 - \theta + \theta^2)$, or 0.125, to the overall LOD score. Family 1415 is similar, but includes only one unaffected sibling of genotype 12.

complete linkage, we can derive the parental phases from the genotypes of the affected offspring. This reveals that 19 of the 26 parental chromosomes carrying the CF allele also carry allele 1 at the *met* locus, while only 11 of 26 parental chromosomes carrying the normal allele at the CF locus carry allele 1 at the *met* locus, which is marginally significant ($\chi^2_1 = 3.86$, $p = 0.05$). The allelic distribution at the *met* locus among the latter chromosomes does not differ significantly from that observed among 60 chromosomes tested in a random Utah control panel ($\chi^2_1 = 0.98$). When these chromosomes carrying a normal allele at the CF locus are pooled and contrasted to the chromosomes carrying the CF deleterious allele, heterogeneity with respect to the *met* alleles 1 and 2 is no longer significant ($\chi^2_1 = 2.71$). Judgment regarding possible allelic association must await the characterization of additional CF families.

The apparent close genetic linkage found in family studies suggests an upper limit of 5% on the genetic distance between the CF locus and the *met* locus. The absence of recombinants between the CF locus and the *met* locus in our study does not support the hypothesis that cystic fibrosis can be caused by mutations at more than one locus. However, ours is still a limited sample and it is important to test many more families to place a more restrictive upper limit on the possible frequency of cystic fibrosis caused by mutations at other loci.

Although the *met* locus is potentially useful as a prenatal diagnostic marker for cystic fibrosis in families known to be at risk, several additional data must be developed before such application. First, more CF families with large sibships must be tested, to refine the genetic distance as well as to detect locus heterogeneity. Observation of a subset of families showing no linkage to the *met* locus could reveal locus heterogeneity. These data will be essential for providing quantitative estimates of

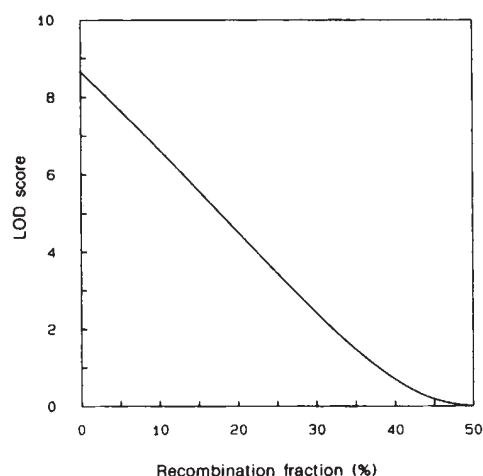


Fig. 2 LOD score for linkage between cystic fibrosis and the *met* locus as a function of recombination fraction. A recessive model with complete penetrance was assumed for cystic fibrosis, with a frequency of 0.025 for the deleterious allele. The results were unaltered over a wide range of allelic frequencies. The test marker was treated as a co-dominant, two-allele system, with gene frequencies of 0.56 and 0.44 for alleles 1 and 2, respectively. The *TaqI* polymorphism was scored as the primary genetic marker and was corroborated by the *MspI* polymorphism. The LOD score is computed as the decimal logarithm of the probability of the observations for a given value of the recombination fraction divided by the corresponding probability under the assumption of no linkage. This probability ratio measures our degree of belief in the hypothesis of linkage at the assumed value relative to the hypothesis of independent segregation. Rather than a standard error based on large sample theory, the use of an empirical confidence region based on the observed LOD score distribution has been recommended¹⁰. A one-unit support limit is obtained by finding that value of the recombination fraction for which the LOD score is decreased by one unit, defining a range of the recombination value for which the odds remain within 10 to 1 of the maximum. This yields an empirical bound for the recombination fraction of 0.051 in our sample. Values for recombinant fraction θ of 0, 0.01, 0.05, 0.10, 0.20, 0.30 and 0.40 produced respectively values for $z(\theta)$ of 8.65, 8.46, 7.66, 6.64, 4.52, 2.41 and 0.70.

individual risk. Second, the extent of polymorphism must be improved so that the marker locus has more than two alleles. At present, many individuals at risk would be unable to obtain useful information. Third, if, as we expect, some recombinants between the *met* and CF loci are found, it will be important to develop a genetic marker locus on the other side of the CF locus to detect recombination events in the region. Until these conditions are met, diagnostic applications are premature.

The fact that no recombinants between the *met* gene locus and the CF locus have yet been identified suggests that the two are indeed very close and even raises speculation that the *met* gene might actually be the cystic fibrosis gene. Furthermore, because the product of the *met* gene is a membrane-spanning receptor protein showing tyrosine kinase homology¹, characteristics which might also be expected of a gene causing a chloride ion-exchange defect, in retrospect *met* might have been considered a candidate gene for the disease. However, although we cannot exclude the *met* gene as a candidate for the CF gene, aetiological involvement of the *met* gene in cystic fibrosis is unlikely: the close linkage still permits a span of more than 1,000 kb within which to place the CF gene and as many as a hundred genes may be located within so wide a region. It is nonetheless interesting that the *met* gene bears homology with the insulin receptor¹, located on chromosome 19 very near the low-density lipoprotein (LDL) receptor²; this homology may suggest that these are members of a clustered family of genes. By this reasoning, although not itself the CF gene, *met* could

open the way to identifying a family of closely linked genes of which one may be the gene for cystic fibrosis.

The most significant aspect of the localization of the CF gene is the impetus it provides to efforts to identify and clone the gene. New techniques make it feasible to cover the large distances that are likely to be involved. Pulse-field gel electrophoresis⁶ and new, rapid 'walking' techniques⁷ suggest approaches to moving closer to the gene. Recent advances in cell culture technology that permit *in vitro* growth of cells which maintain the CF phenotype⁸ suggest the possible invention of protocols that could provide functional assays for regions that might encompass the gene. We hope that identification of the CF gene will permit identification and analysis of the biochemical pathway which is perturbed by mutations in the gene, and will suggest rational means of intervention and management of this disease.

We thank Barbara Ogden for assistance in family collection, Tara Bragg, John Cowan, Russell Eldridge, Mary Hadley, Leslie

Jerominski, Lesa Nelson and Leslie Sargent for technical assistance, Carol Wood of the Intermountain Cystic Fibrosis Center for assistance in family collection, the NIGMS Human Genetic Mutant Cell Repository for lymphoblastoid cell lines of family 774, and Ruth Foltz for editorial assistance in preparing this manuscript. M.D. is an American Cancer Society postdoctoral fellow.

Received 1 November; accepted 6 November 1985.

1. Dean, M. *et al.* *Nature* **318**, 385-388 (1985).
2. Eiberg, H. *et al.* 8th int. Workshop Human Gene Mapping, Helsinki, Abstr. 100 (Karger, New York, in the press).
3. Cooper, C. S. *et al.* *Nature* **311**, 29-33 (1984).
4. Lathrop, G. M. & Lalouel, J.-M. *Am. Soc. hum. Genet.* **36**, 460-465 (1984).
5. 8th int. Workshop Human Gene Mapping, Helsinki, draft rep. 223 (Karger, New York, in the press).
6. Schwartz, D. & Cantor, C. *Cell* **37**, 67-75 (1984).
7. Collins, F. & Weissman, S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6812-6816 (1984).
8. Stutts, M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **82**, 6677-6681 (1985).
9. Barker, D., Holm, T. & White, R. *Am. J. hum. Genet.* **36**, 1159-1171 (1984).
10. Conneally, P. *et al.* 8th Int. Workshop Human Gene Mapping, Helsinki (Karger, New York, in the press).

Localization of cystic fibrosis locus to human chromosome 7cen-q22

Brandon J. Wainwright*, Peter J. Scambler*, Jorg Schmidtke†, Ella A. Watson*, Hai-Yang Law*, Martin Farrall*, Howard J. Cooke‡, Hans Eiberg§ & Robert Williamson*

* Department of Biochemistry, St Mary's Hospital Medical School, University of London, London W2 1PG, UK

† Institut für Humangenetik der Universität, Gosslerstrasse 12d, D-3400 Göttingen, FRG

‡ MRC Mammalian Genome Unit, West Mains Road, Edinburgh EH9 3JJ, UK

§ Institute of Medical Genetics, Panum Institute, Copenhagen, Denmark

Cystic fibrosis (CF) is the most common genetic disease in Caucasian populations, with an incidence of 1 in 2,000 live births in the United Kingdom, and a carrier frequency of approximately 1 in 20. The biochemical basis of the disease is not known¹, although membrane transport phenomena associated with CF have been described recently². Consanguinity studies have shown that the inheritance of CF is consistent with it being a recessive defect caused by a mutation at a single autosomal locus³. Eiberg *et al.*⁴ have reported a genetic linkage between the CF locus and a polymorphic locus controlling activity of the serum aryl esterase paraoxonase (PON). The chromosomal location of PON, however, is not known⁵. Linkage to a DNA probe, DOCR1-917, was also recently found at a genetic distance of ~15 centimorgans (L.-C. Tsui and H. Donniss-Keller, personal communication), but no chromosomal localization was given. Here we report tight linkage between the CF locus and an anonymous DNA probe, pJ3.11, which has been assigned to chromosome 7cen-q22.

The use of DNA probes allows the determination of linkage between a restriction fragment length polymorphism (RFLP) and a mutated locus causing pathology, even when the biochemical defect is unknown. We have used this approach successfully to analyse sex-linked diseases such as Duchenne mus-

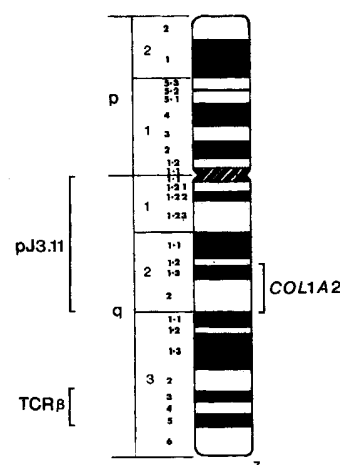


Fig. 1 Map of human chromosome 7, showing the chromosomal locations of probes *TCRβ* (ref. 9), *COL1A2* (ref. 25) and pJ3.11 (ref. 18).

cular dystrophy⁵; it has also been used to study previously unassigned autosomal dominant diseases such as Huntington's chorea⁶ and adult onset polycystic kidney disease⁷. However, although linkage has been attempted with DNA markers to autosomal recessive diseases of unknown biochemical aetiology, only exclusions have been achieved to date.

There have been few clues to the chromosomal localization of the CF locus. Mayo *et al.*⁸ suggested that the CF locus might be localized on chromosome 4, based on data from cell hybrids expressing a ciliary dyskinesia factor, but this was subsequently shown to be unlikely by exclusion mapping with several DNA and protein markers⁹. Two different families have been reported to show both chromosomal anomalies and CF; the study of one of these families suggested that the CF locus mapped to 5p, whereas in the other family it was mapped to 13q34 (ref. 10). We investigated the latter family by mapping the gene for clotting factor X to 13q34 (ref. 11) and then excluded CF from linkage to factor X¹². We have also excluded the candidate gene coding for complement component 3 by showing that alleles for an

Table 1 LOD scores at various recombination fractions (θ) for the relationships between the CF locus (CF) and the locus defined by pJ3.11 and between CF and *TCRβ*

Marker	No. of meioses	$\theta = 0$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
CF versus J3.11	27	5.24	4.52	3.80	3.09	2.41	1.77	1.20	0.77	0.33	0.11
TCRβ versus CF	22	—∞	1.87	2.06	1.87	1.54	1.17	0.80	0.47	0.22	0.06

Combined LOD scores were calculated using the computer program package LINKAGE as described by Lathrop²⁴.

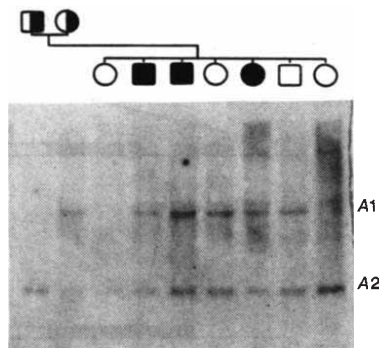


Fig. 2 Segregation of the *MspI* polymorphism defined by pJ3.11 with the CF locus in a family with three affected (solid symbols) and three unaffected (open symbols) children. The RFLP alleles are 4.2 kilobases (kb) (A1) and 1.8 kb (A2) long; the frequency of the rare allele A1 is 0.2 (ref. 26).

Methods. DNA (5 μ g) from 10 informative families with two or three affected children was digested with *MspI* and fractionated by electrophoresis on 0.8% agarose gels by standard methods⁹. Fractionated DNA was transferred to Hybond membranes (Amersham International) and hybridized⁹. The probe used was the 500-bp *HindIII*-*EcoRI* fragment from the insert of pJ3.11, labelled to a specific activity of 1×10^9 d.p.m. μ g⁻¹ by synthesis using random oligonucleotide primers. Blots were washed to a final salt concentration of 0.15 M NaCl in the presence of 0.2% SDS. Autoradiography was for 16 h at -70° using double-intensification screens.

RFLP detected by this sequence do not co-segregate with cystic fibrosis in families with multiple affected sibs¹³. Several groups have reported extensive linkage exclusions between the CF locus and protein and DNA markers¹⁴⁻¹⁶, which in total exclude ~40% of the human genome.

We have tested for linkage between the CF locus and multiple markers on a single human chromosome by multipoint linkage analysis¹⁷. As part of our attempt to study human chromosomes for which few exclusions had been obtained for cystic fibrosis, we have examined the linkage relationships of the two DNA markers located on chromosome 7 to the CF locus. These are the anonymous DNA segment pJ3.11, which defines an *MspI* polymorphism and has been mapped to 7cen-q22 by somatic cell hybrid panels¹⁸, and the T-cell receptor β -chain (*TCR\beta*) gene which defines *BglII* polymorphism and maps to 7q3 (ref. 19). These localizations are summarized in Fig. 1.

The linkage relationships between the two markers and the CF locus are shown in Table 1. The maximal LOD score between the locus defined by pJ3.11 and the CF locus calculated for combined sexes is 5.24 at a recombination fraction (θ) = 0 (99% confidence interval 0-13 centimorgans). A LOD score of 3.0 is accepted as a sufficient statistic to prove synteny between two loci²⁰. No recombinants between the probe and the CF locus have been observed in 27 meioses. A typical pedigree is shown in Fig. 2. All informative pedigrees and typings are available from the authors.

TCR\beta (7q3) shows significant linkage (for two syntenic loci²¹) to the CF locus at a genetic distance of 10 centimorgans. We have also demonstrated that the collagen locus *COL1A2* is linked to the CF locus at a genetic distance of ~10 centimorgans²², suggesting that *COL1A2* (7q22) and *TCR\beta* (7q3) may flank the CF locus, as they are not themselves closely linked. We are presently analysing the linkage between pJ3.11, *TCR\beta* and *COL1A2* in large reference pedigrees (Centre d'Etude Polymorphisme Humain), as well as in cystic fibrosis families which have been typed for *PON*.

Genetic heterogeneity of the locus mutated in cystic fibrosis (as opposed to multiple mutations causing a defect at a single locus) would be demonstrated by the observation of heterogeneity in the linkage relationship in different families between a given close marker and the CF locus. The informative families in this study originated from Australia, Eire, France,

West Germany, United Kingdom and Pakistan. All the families show tight linkage between the pJ3.11 probe and the CF locus. It would be of interest to study these probes in the Amish kindred reported by Klinger²³, as there are data indicating differences in linkage for that family compared with those studied here (K. Klinger, personal communication). A more extensive analysis covering different populations is in progress.

The pJ3.11 probe is presently informative in approximately 40% of matings and, given the tight linkage demonstrated here, would provide accurate heterozygote detection and antenatal diagnosis of cystic fibrosis in some cases. However, such diagnoses will be facilitated markedly when the availability of a set of informative and ordered flanking markers (*PON*, pJ3.11, *COL1A2*, and *TCR\beta*) has been confirmed.

We thank the Cystic Fibrosis Research Trust, and especially Ron Tucker, the MRC and the Foulkes Foundation for consistent and generous financial support, and our many clinical colleagues and the CF families for their help in obtaining and giving samples. We particularly thank Kay Davies, Lap-Chee Tsui, Kathy Klinger, Veronica van Heyningen and Nigel Spurr for helpful discussions, and for providing data in advance of publication.

Received 4 November; accepted 7 November 1985.

- Williamson, R. et al. in *Cystic Fibrosis—Horizons* (ed. Lawson, D.) 139-153 (Wiley, Chichester, 1984).
- Quinton, P. M. & Bijman, J. *New Engl. J. Med.* **308**, 1185 (1983).
- Romeo, G. et al. *Am. J. hum. Genet.* **37**, 334-338 (1985).
- Eiberg, H., Mohr, J., Schmiegelow, K., Nielsen, L. S. & Williamson, R. *Clin. Genet.* **28**, 265-271 (1985).
- Davies, K. E. et al. *Nucleic Acids Res.* **11**, 2303-2312 (1983).
- Gusella, J. F. et al. *Nature* **306**, 234-238 (1983).
- Reeders, S. T. et al. *Nature* **317**, 542-544 (1985).
- Mayo, B. J., Klebe, R. J., Barnett, D. R., Lankford, B. J. & Bowman, B. H. *Clin. Genet.* **18**, 329-386 (1980).
- Scambler, P., Gilliam, C., Tippet, P., Boylston, A. & Williamson, R. *Hum. Genet.* **82**, 6167-6171 (1985).
- Edwards, J. H., Jonasson, J. A. & Blackwell, N. L. *Lancet* **i**, 1020 (1984).
- Scambler, P. J. & Williamson, R. *Cytogenet. Cell Genet.* **39**, 231-233 (1985).
- Scambler, P. J., Wainwright, B. J., MacGillivray, R., Fung, M. R. & Williamson, R. *Am. J. hum. Genet.* (in the press).
- Davies, K. E., Gilliam, T. C. & Williamson, R. *Molec. Biol. Med.* **1**, 185-190 (1983).
- Scambler, P. J. et al. *Cytogenet. Cell Genet.* (in the press).
- Tsui, L.-C., Kennedy, D., Zsiga, M. & Buchwald, M. *Am. J. hum. Genet.* **36**, 156 (1984).
- Bowcock, A. M., Lee, G. M., Sanders, M. F. & Cavalli-Sforza, L. *8th int. Workshop Human Gene Mapping*, Helsinki, abstr. 37 (Karger, New York, in the press).
- Farrall, M., Scambler, P., North, P. & Williamson, R. *Am. J. hum. Genet.* (in the press).
- Bartels, I., Grzeschik, K.-H., Cooper, D. & Schmidtke, J. *Am. J. hum. Genet.* (in the press).
- Morton, C. C. et al. *8th int. Workshop Human Gene Mapping*, Helsinki, abstr. 256 (Karger, New York, in the press).
- Morton, N. *Am. J. hum. Genet.* **7**, 277-318 (1955).
- Chotai, J. *Ann. hum. Genet.* **48**, 359-378 (1984).
- Scambler, P. J. et al. *Lancet* (submitted).
- Klinger, K. *Hum. Genet.* **65**, 94-98 (1983).
- Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3443-3446 (1984).
- Tsipouras, P. et al. *Am. J. hum. Genet.* **36**, 1172-1179 (1984).
- Cooper, D. N., Smith, B. A., Cooke, H. J., Niemann, S. & Schmidtke, J. *Hum. Genet.* **69**, 201-205 (1985).

The human *met* oncogene is related to the tyrosine kinase oncogenes

Michael Dean*, Morag Park†, Michelle M. Le Beau‡, Terry S. Robins*, Manuel O. Diaz‡, Janet D. Rowley‡, Donald G. Blair† & George F. Vande Woude*

* LBI-Basic Research Program, NCI-Frederick Cancer Research Facility and † Laboratory of Molecular Oncology, National Cancer Institute, PO Box B, Frederick, Maryland 21701, USA

‡ Joint Section of Hematology/Oncology, University of Chicago, Chicago, Illinois 60637, USA

The *met* oncogene was previously isolated from a chemically transformed human cell line, MNNG-HOS (refs 1, 2). Recent evidence has demonstrated that two classes of transcripts are expressed from the *met* proto-oncogene locus. The *met* oncogene, however, expresses an aberrant RNA which has sequences in common with both transcripts. We now report partial nucleotide sequencing of the human *met* oncogene and show that *met* is

related to the protein kinase oncogenes and growth factor receptors. The *met* nucleotide sequence is not identical to that of any published gene, and it is more closely homologous to the tyrosine kinases than to the serine/threonine kinases. Within the tyrosine kinase family, the sequenced *met* domains are most closely related to the human insulin receptor and the viral *abl* gene. *In situ* chromosome hybridization has mapped *met* to human chromosome 7 band 7q21-q31, a location distinct from that of other kinases. This is also a region associated with nonrandom chromosomal deletions observed in a portion of patients with acute non-lymphocytic leukaemia. The accompanying paper³ shows that this chromosomal locus is also tightly linked with the human hereditary disease cystic fibrosis.

To characterize the *met* oncogene and its protein product(s), we have sought to determine the nucleotide sequence of coding portions of the locus. The sequences required for transformation by *met* are contained within 35 kilobases (kb) of human genomic DNA and NIH 3T3 cells transformed with *met* express a 6.5-kb messenger RNA^{1,4}. Initial attempts to isolate a complementary DNA copy of this RNA generated a 2.0-kb clone. However, the partial nucleotide sequence indicated that this clone contains only untranslated sequences (M.P., data not shown). As an alternative approach, a 9-kb fragment was cloned into a retrovirus vector derived from Moloney sarcoma virus (Fig. 1).

Retroviral vectors have been shown to process segments of genomic DNA, generating cDNA-like copies⁵. We hoped that by cloning a segment of *met* genomic DNA into such a vector we could recover a clone devoid of intervening sequences. An *S*₁ nuclease experiment revealed that passage of this segment of *met* genomic DNA through a retroviral intermediate resulted in the excision of at least one intervening sequence; the recovered clone protected a fragment 1–200 bases longer than the corresponding segment of genomic DNA (data not shown). Nucleotide sequencing of the recovered clone and the corresponding portions of genomic DNA revealed an open reading frame 126 amino acids long (Fig. 2A). When this putative *met* amino-acid sequence was used to search an amino-acid sequence database, homology with several tyrosine kinase oncogenes and growth factor receptors was revealed (Fig. 2B).

In addition, we have also determined the nucleotide sequence of a 1.1-kb fragment of genomic DNA from the middle of the *met* oncogene (fragment D) which is known to hybridize with *met* mRNA transcripts². This sequence contained an open reading frame bounded by splice acceptor and donor signals which shows dramatic homology to the tyrosine kinases (Fig. 2). Figure 2B compares the amino-acid sequence of the two open reading frames with several members of the tyrosine kinase family. The putative exon is identical in 18/23 residues with the human insulin receptor^{6,7} and in 15/23 amino acids with the virus *abl*

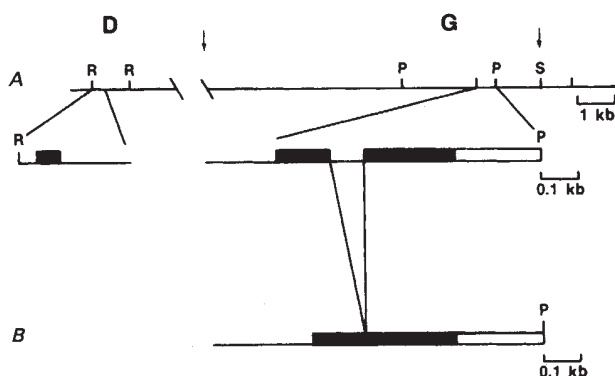


Fig. 1 Sequencing strategy of the *met* oncogene. **A**, Genomic DNA located at the 3' end. Filled areas depict coding regions or putative exons. **D** and **G** refer to fragments described previously^{1,2}. Both strands of the expanded regions were sequenced by the chain-termination method²⁶. The 9-kb *SalI* fragment inserted into the retrovirus vector is depicted by the arrows. **B**, *met* sequences contained within the spliced clone recovered. R, *EcoRI*; P, *PstI*; S, *SalI*.

Methods. The 9-kb *SalI* fragment was cloned into pGV16 (ref. 27), which was modified to accept *SalI* fragments by inserting a *SalI* linker. Virus was rescued from this plasmid by co-transfection with an equal amount of pMOV-3 plasmid²⁸ into NIH 3T3 cells. Infected NIH 3T3 cells were subsequently selected for resistance to G418 using 500 $\mu\text{g ml}^{-1}$ Geneticin (Gibco). The viral DNA from the transfected cells was recombined by fusing G418-resistant NIH 3T3 cells with COS cells using polyethylene glycol as described previously²⁹. Hirt supernatants³⁰ were prepared after 72 h and used to transform *Escherichia coli* strain HB101 (BRL) using the procedure described by Hanahan³¹ and selected on kanamycin (35 $\mu\text{g ml}^{-1}$). Only colonies which hybridized to the G fragment probe (see text) were analysed.

gene⁸. This region is highly conserved among all members of the tyrosine kinase family⁹, and moderately conserved between other kinases such as *mos* (ref. 10) (Fig. 2B), *raf* (ref. 11), the yeast CDC28 protein¹² and cyclic AMP-dependent kinase¹³. Although the two exons at the C-terminus of *met* contain a region which is less well conserved among kinases⁹, certain portions of this domain are conserved among all members of the family, including *met* (Fig. 2B).

We have previously demonstrated that although *met* maps to chromosome 7, it is not the *erbB*/epidermal growth factor (EGF) receptor gene, which is located at chromosome 7 p12-p14 (refs 1, 14). To map the location of *met* on chromosome 7 more precisely, we performed *in situ* chromosomal hybridization. A ³H-radiolabelled *met* probe was hybridized to metaphase cells

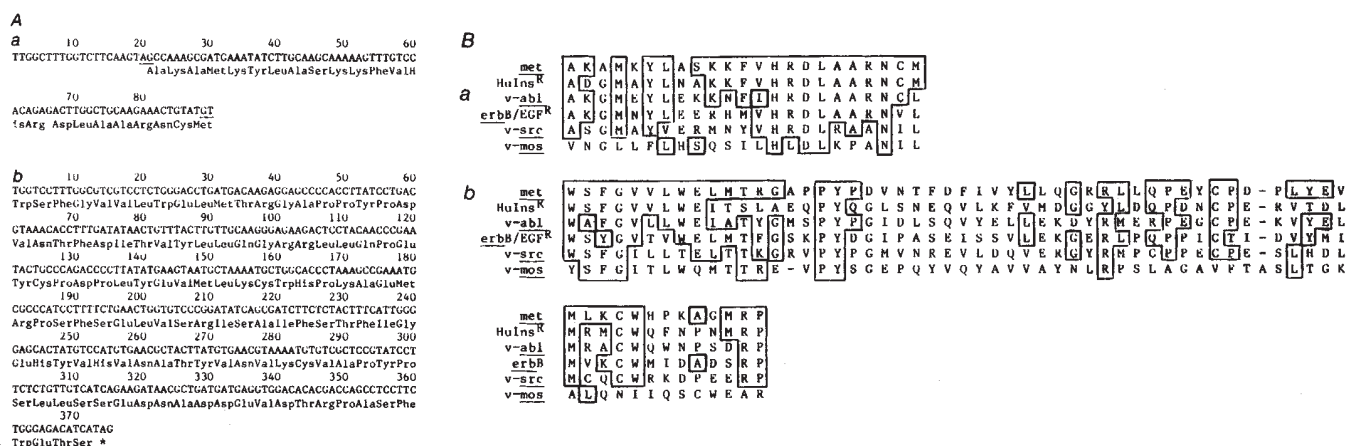


Fig. 2 Partial nucleotide sequence of the *met* oncogene. **A**, The nucleotide sequence and inferred amino-acid sequence of portions of *met* are shown: **a**, putative exon from D fragment; **b**, open reading frame of the last two exons. Putative splice signals are underlined. **B**, Comparison of the *met* amino-acid sequence with corresponding regions of the human insulin receptor (Hu InsR)^{6,7}, viral *abl* oncogene⁸, human EGF receptor³², and viral *src* and human *mos* proto-oncogenes^{10,33}.

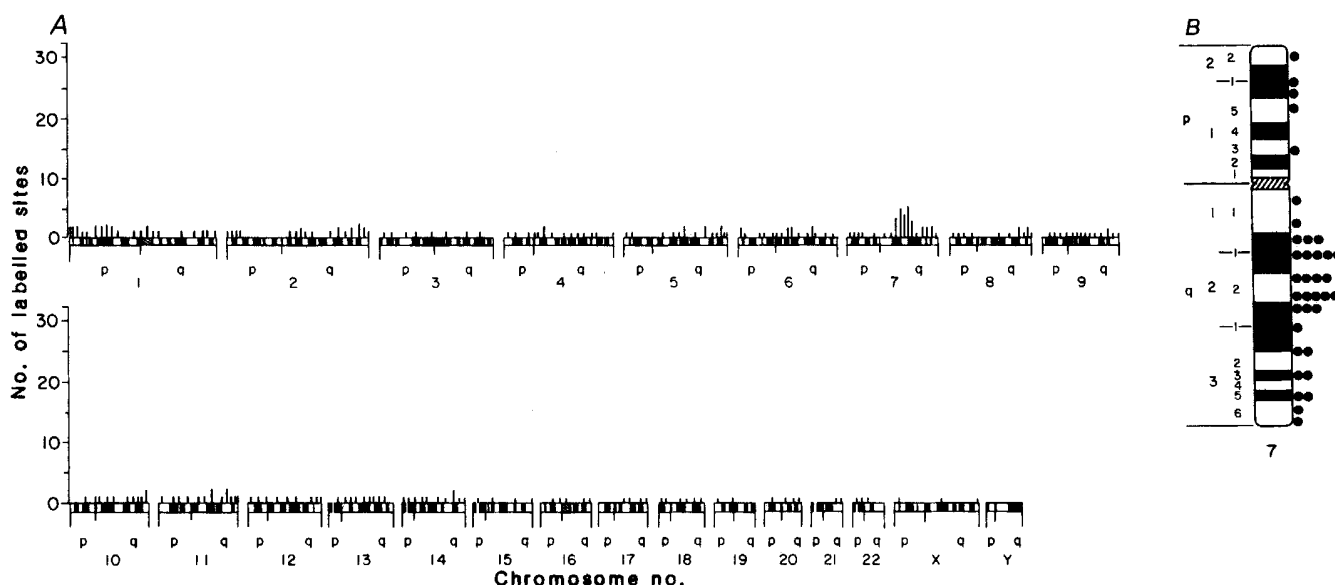


Fig. 3 *In situ* chromosomal hybridization of the *met* oncogene. **A**, Distribution of labelled sites in 100 normal metaphase cells. Human metaphase cells prepared from phytohaemagglutinin-stimulated peripheral blood lymphocytes were hybridized with the ^3H -labelled E probe¹. Specific labelling was observed on chromosome 7 at bands q21–q31. **B**, Distribution of labelled sites on chromosome 7. Of 100 metaphase cells examined, 19 (19%) were labelled on bands q21–q31 of chromosome 7. Of a total of 36 grains on chromosome 7, 21 or 58% were located on q21–q31; these sites represented 8.2% (21/256) of all labelled sites ($P > 0.005$). A modification of the method of Harper and Saunders was used^{34,35}.

from phytohaemagglutinin-stimulated peripheral blood lymphocytes. Figure 3A shows the distribution of grains observed in 100 metaphase cells. A significant clustering of grains was observed on the long arm of chromosome 7 at bands q21–q31 ($P < 0.005$); no other chromosome showed specific labelling. Of the 100 metaphase cells examined, 19 were labelled on bands q21–q31 of one or both of the chromosome 7 homologues (Fig. 3B); these sites represented 8.2% (21/256) of all labelled sites. The largest cluster of grains was observed at band 7q22. Specific labelling was also observed at this chromosomal region when *met* was hybridized to metaphase cells from peripheral blood lymphocytes obtained from a second individual. Of 50 metaphase cells examined, 10 showed label on one or both chromosome 7 homologues, representing 7.1% (12/169) of all labelled sites. These results confirm the location of this gene on the long arm of human chromosome 7 and refine the location to band q21–q31.

We conclude that the *met* oncogene is a member of the tyrosine kinase family, although analysis of the *met* protein product will be required to establish whether it possesses this activity. Analysis of *met*-related mRNAs has revealed that the transforming gene expresses a hybrid truncated mRNA⁴. Thus, we speculate that the *met* oncogene transforms cells by expressing a truncated form of the kinase encoded by the proto-oncogene. A similar mechanism appears to be responsible for the activation of both the *v-erb-B* and *c-erb-B*/EGF receptor genes^{15,16}, and the *bcr/abl* gene in chronic myelogenous leukaemia¹⁷.

The homology between *met* and the kinase domain of the insulin receptor is intriguing. Insulin is a member of a gene family which includes insulin-like growth factors I and II and nerve growth factor¹⁸. Similarly, there are several distinct receptors for insulin-related mitogens¹⁹. It has not escaped our attention that the *met* class I gene product may encode a receptor for an insulin-like peptide. However, *met* is also closely related to *abl*. Identification and characterization of the *met* protein(s) should help address these interesting questions.

Nonrandom deletions of the long arm of chromosome 7 are frequently observed in the leukaemic cells of patients with acute non-lymphocytic leukaemia (ANLL). Interestingly, *met* was isolated from a human cell line (HOS) transformed with the chemical carcinogen *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine, and increasing evidence now relates the development of ANLL

to mutagen exposure both in patients subjected to cytotoxic therapy for a primary malignancy and in patients exposed to environmental carcinogens^{20,21}.

It is notable that recurring abnormalities involving chromosome 7 are common in such patients. These abnormalities include loss of a whole chromosome 7 or of part of the long arm of this chromosome, *del*(7q) (refs 20, 22). It may be significant that *met* is located in the region of the proximal breakpoint of the interstitial deletions of 7q observed in patients with ANLL²³.

In the accompanying paper, White *et al.*³ describe restriction fragment length polymorphisms (RFLPs) from the *met* locus which were developed for examining ANLL patients. They find that these RFLPs are tightly linked to the hereditary disease cystic fibrosis (CF). The *in situ* chromosomal mapping of *met* thus localizes the CF locus in man. At the very least then, *met* is a marker for CF, a disorder that is speculated to involve an ion-channel imbalance²⁴. However, we show here by sequence homology that *met* is a member of the tyrosine kinase family and we note that tyrosine kinase receptor activation has been linked to the regulation of the sodium ion-proton antiport²⁵. We also note that the *met* proto-oncogene locus expresses two transcripts which appear to be noncoordinately expressed overlapping transcripts³. The genomic sequences encoding these two transcripts could cover greater than several hundred kilobases. Present genetic data³ suggest that the distance between *met* and the CF locus is less than 5,000 kb.

We thank Doug Ludtke and Martin Schmidt for assistance with DNA sequencing, Chris Westphal and Linda Miller for technical assistance, and Friedrich Propst for help with the computer analysis. Research was sponsored by the National Cancer Institute, DHHS, under contract NO1-CO-23909 with Litton Bionetics, Inc. Additional support was provided by contract DE-ACO2-80EV 10360 from the US Department of Energy, PHS grant CA 23954 from the NCI and grant 84-20 from the Illinois Division of the American Cancer Society. M.D. is a Fellow of the ACS and M.M.L.B. is a Special Fellow of the Leukemia Society of America. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US government.

Received 31 October; accepted 6 November 1985.

- Cooper, J. A. *et al. Nature* **311**, 29-33 (1984).
- Cooper, C. S. *et al. Cancer Res.* **44**, 1-10 (1984).
- White, R. *et al. Nature* **318**, 382-384 (1985).
- Park, M. *et al.* (in preparation).
- Robins, T., Jhappan, C., Chirikjian, J. & Vande Woude, G. F. *Gene analysis Techq.* (in the press).
- Ullrich, A. *et al. Nature* **313**, 756-761 (1985).
- Ebina, Y. *et al. Cell* **40**, 747-758 (1985).
- Reddy, E. P., Smith, M. J. & Srinivasan, A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3623-3627 (1983).
- Hunter, T. & Cooper, J. A. in *The Enzymes* (eds Boyer, P. D. & Krebs, E.) (Academic, New York, in the press).
- van Beveran, D., van Straaten, F., Gallegher, J. A. & Verma, I. M. *Cell* **27**, 97-108 (1981).
- Bonner, T. I. *et al. Molec. cell. Biol.* **5**, 1400-1407 (1985).
- Lorinez, A. T. & Reed, S. I. *Nature* **307**, 183-185 (1984).
- Shoji, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 848-851 (1981).
- Merlino, G. T. *et al. Molec. cell. Biol.* **5**, 1722-1734 (1985).
- Yamamoto, T. *et al. Cell* **35**, 71-78 (1983).
- Nilsen, T. W. *et al. Cell* **41**, 719-726 (1985).
- Shitelman, E., Lifshitz, B., Gale, R. P. & Canaan, E. *Nature* **315**, 550-553 (1985).
- Froesch, E. R., Schmid, C., Shwander, J. & Zapf, J. *A. Rev. Physiol.* **47**, 443-467 (1985).
- Rechler, M. M. & Nissley, S. P. *A. Rev. Physiol.* **47**, 426-442 (1985).
- Rowley, J. D., Golomb, H. M. & Vardiman, J. W. *Blood* **58**, 759-767 (1981).
- Mitelman, F. *et al. Cancer Genet. Cytogenet.* **9**, 197-214 (1981).
- Yunis, J. J. *Cancer Genet. Cytogenet.* **11**, 125-137 (1984).
- Rowley, J. D. & Testa, J. R. *Adv. Cancer Res.* **36**, 103-149 (1982).
- van Heyningen, V. *Nature* **311**, 104-105 (1984).
- l'Allemain, G., Paris, S. & Pouyssegur, J. *J. biol. Chem.* **259**, 5809-5815 (1984).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Jhappan, C., Vande Woude, G. & Robins, T. *J. Virol.* (submitted).
- Harbers, K., Schnieke, A., Stuhlman, H., Jahner, D. & Jaenisch, R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7609-7613 (1981).
- Pontecorvo, G. *Somat. Cell Genet.* **1**, 397-400 (1975).
- Hirt, B. *J. molec. Biol.* **26**, 365-369 (1967).
- Hanahan, D. *J. molec. Biol.* **166**, 557-586 (1983).
- Downward, J. *et al. Nature* **307**, 521-527 (1984).
- Czernilofsky, A. P. *et al. Nature* **287**, 198-203 (1980).
- Harper, M. E. & Saunders, G. F. *Chromosoma* **83**, 431-439 (1981).
- LeBeau, M. M., Westbrook, C., Diaz, M. O. & Rowley, J. D. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6692-6696 (1985).

Retroviral gag and DNA endonuclease coding sequences in IgE-binding factor gene

Hiroyuki Toh*, Masao Ono† & Takashi Miyata*‡

* Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan

† Department of Molecular Biology, School of Medicine, Kitasato University, Sagamihara-shi, Kanagawa 228, Japan

‡ To whom correspondence should be addressed.

Immunoglobulin-binding factors are known to regulate the synthesis of B-cell-derived immunoglobulin heavy-chain isotypes¹. Cloning and nucleotide sequence determination of complementary DNA encoding rodent IgE-binding factors (IgE-BF) revealed that messenger RNA encodes a glycoprotein of 557 amino acids which is expressed as a precursor of relative molecular mass (M_r) 60,000 (60K) in COS7 monkey cells². We report here that the 3' two-thirds of the IgE-BF coding sequence shows a surprising homology (72%) at the DNA level with coding sequences of the *gag* and *pol* (DNA endonuclease) genes of the Syrian hamster intracisternal A particle (IAP H18)³, an endogenous retrovirus. This marked homology demonstrates that the rodent gene encoding IgE-BF is a hybrid gene which evolved very recently by integrating genes of viral origin, and that the encoded polypeptide comprises three separate domains: an IgE-BF domain and retrovirus-derived *gag* and DNA endonuclease-like domains. This may represent the first report of a cellular gene containing a virus-derived coding sequence.

Figure 1 shows a homology matrix comparison between the nucleotide sequence coding for the rodent IgE-BF and that of

the Syrian hamster IAP genome³. A striking homology with IAP is found for the 3' two-thirds (nucleotides 700-1,770) of the IgE-BF coding sequence, whereas the remaining region (1-699) shares no obvious homology with IAP. In IAP, the homology extends over two regions (nucleotides 1,483-2,140 and 5,364-5,772) which are interrupted by a non-homologous sequence of ~3,200 nucleotides bounded by direct repeats of non-identical but homologous sequences (2,142-2,185 and 5,364-5,406). This non-homologous sequence is completely absent in the IgE-BF coding sequence (Fig. 1b). The direct repeats may result in the deletion of a large region of DNA in the IgE-BF coding sequence. The 5' homologous region of IAP exists within the *gag* coding region, while the 3' block contains part of the DNA endonuclease coding sequence^{3,4}. Alignment (Fig. 2) of the nucleotide sequences reveals 72% homology at the DNA level. Martens *et al.*² noted that the deduced amino-acid sequence of the IgE-BF precursor polypeptide shows local homology with retroviral *pol* gene products in the C-terminal region of 50 amino acids. This high degree of sequence homology indicates unequivocally that the two sequences are evolutionarily related.

These results strongly suggest that the rodent IgE-BF precursor comprises two components of distinct origin: the N-terminal one-third (amino-acid positions 1-202) carries a function unique to IgE-BF while the remaining C-terminal portion consisting of *gag* (203-422) and DNA endonuclease (423-557) related domains was derived from a retrovirus or endogenous retrovirus very recently during evolution. That the IgE-BF coding sequence shows a high degree of nucleotide sequence homology with the rodent IAP, together with the existence of *pol*-related sequences in various retroviruses and transposons across a wide taxonomic distance⁴⁻⁹, strongly suggests that a certain type of retrovirus or endogenous retrovirus was inserted into a cellular DNA region proximal to the primordial IgE-BF gene, and these were later

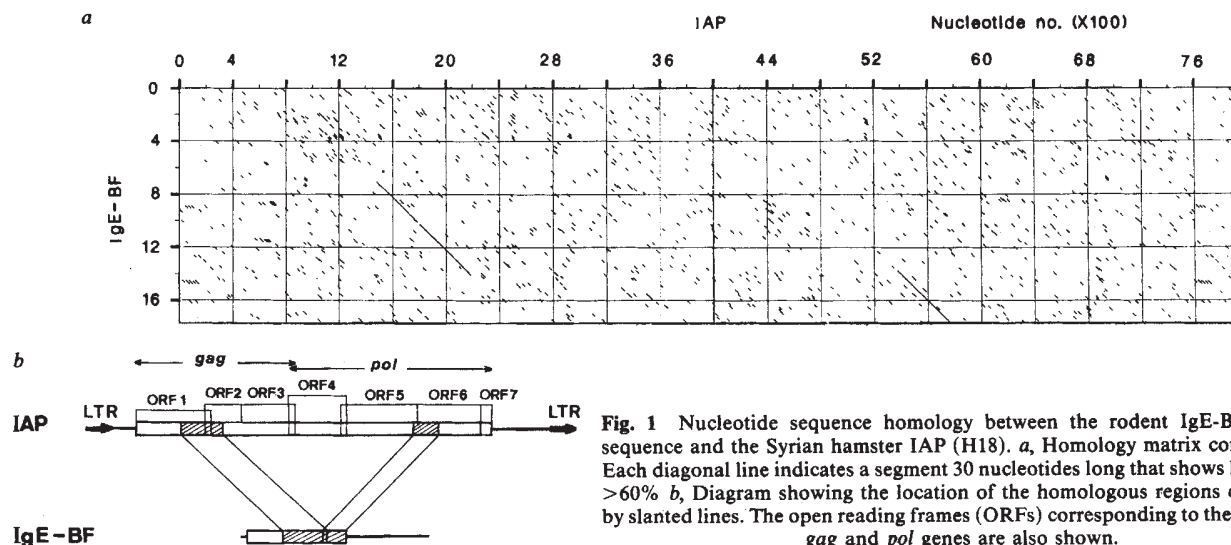


Fig. 1 Nucleotide sequence homology between the rodent IgE-BF coding sequence and the Syrian hamster IAP (H18). a, Homology matrix comparison. Each diagonal line indicates a segment 30 nucleotides long that shows homology >60%. b, Diagram showing the location of the homologous regions connected by slanted lines. The open reading frames (ORFs) corresponding to the retroviral *gag* and *pol* genes are also shown.

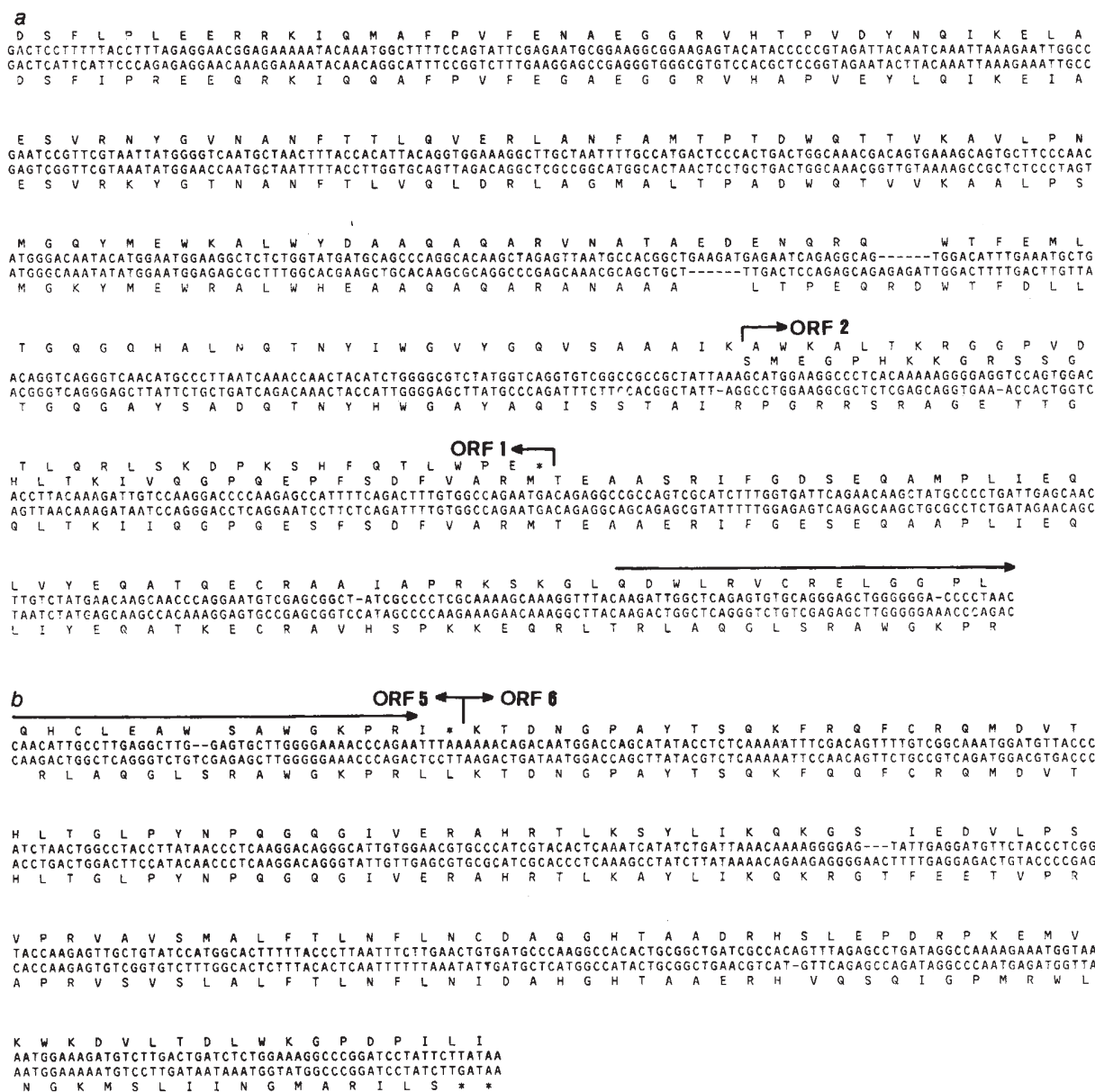


Fig. 2 Alignment of the nucleotide sequences of IAP (upper sequence) and the IgE-BF coding region (lower sequence). **a** and **b** correspond to the 5' and 3' homologous blocks of IAP indicated in Fig. 1. The encoded amino-acid sequences are shown above and below the nucleotide sequences. Direct repeats in IAP are indicated by horizontal arrows. Bars indicate gaps. * Termination codon.

fused to each other. The IgE-BF contains a long stretch of 3'-noncoding sequence². It is expected that this region also shows sequence homology with IAP. It would be of interest to know whether or not IgE-BF genes from non-rodent mammals also contain a similar sequence in their 3' portion.

This may represent the first report of cellular genes that can integrate viral genes. There are several lines of evidence for the reverse case that viruses can integrate cellular genes into their own genomes¹⁰⁻¹⁴. Such virus-mediated gene transfer would allow cellular genes to be shuffled, leading to the emergence of a new gene with a novel function.

Martens *et al.*² showed that the IgE-BF expressed in COS7 cells are glycoproteins of M_r 60,000 and 11,000. As the putative protein-coding sequence of a cloned cDNA predicts a 62K protein, Martens *et al.* suggested that the smaller species is a product derived from the 60K precursor by proteolytic cleavage. Two pairs of potential cleavage sites that would yield polypeptides of ~105 amino acids each and containing an N-glycosylation site were also suggested: Arg-Lys(amino acids 246-247) and Arg-Arg(350-351) or Arg-Lys-Arg(104-106) and Arg-

Lys(212-213)². Given our results, the latter pair is more likely to delineate the 11K protein than the former.

We thank Dr H. Hayashida for discussion and Miss K. Mitsuyasu for technical assistance. This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan.

Received 5 August; accepted 4 October 1985.

1. Ishizaka, K. *A. Rev. Immun.* **2**, 159-182 (1984).
2. Martens, C. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 2460-2464 (1985).
3. Ono, M., Toh, H., Miyata, T. & Awaya, T. *J. Virol.* **55**, 387-394 (1985).
4. Toh, H. *et al. EMBO J.* **4**, 1267-1272 (1985).
5. Toh, H., Hayashida, H. & Miyata, T. *Nature* **305**, 827-829 (1983).
6. Saigo, K. *et al. Nature* **312**, 659-661 (1984).
7. Michel, F. & Lang, B. F. *Nature* **316**, 641-643 (1985).
8. Miyata, T., Toh, H. & Saigo, K. *Proc. Japan Acad.* (in the press).
9. Varmus, H. E. *Nature* **314**, 583-584 (1985).
10. Bishop, J. M. & Varmus, H. E. *RNA Tumour Viruses, Molecular Biology of Tumor Viruses* 2nd edn (eds Weiss, R., Teich, N., Varmus, H. E. & Coffin, J.) 999-1108 (Cold Spring Harbor Laboratory, New York 1982).
11. Kwok, T. J. & Engler, J. A. *Nucleic Acids Res.* **12**, 3959-3971 (1984).
12. Blomquist, M. C., Hunt, L. T. & Barker, W. C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7363-7367 (1984).
13. Brown, J. P., Twardzik, D. R., Marquardt, H. & Todaro, G. J. *Nature* **313**, 491-492 (1985).
14. Baldwin, G. S. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1071-1075 (1985).

A major retroviral core protein related to EPA and TIMP

WE add another curious note to the sequence similarities shared by the proteins identified by biological properties as either the erythroid-potentiating activity (EPA) factor associated with stimulation of erythroid precursors¹ or the tissue inhibitor of metalloproteinases (TIMP). A recent report by Docherty *et al.*² demonstrates that these two proteins have identical sequences.

Using the Intelligenetics PEP program, we have detected regions of similarity between the EPA-TIMP protein and the gag core proteins encoded by retroviruses including human T-cell lymphotropic virus types I, II and III (HTLV-I, -II and -III), bovine leukaemia virus (BLV), Rous sarcoma virus (RSV) and Moloney murine leukaemia virus (Mo-MuLV). These regions are congruent with those conserved among the major retroviral capsid proteins.

These gag proteins are highly divergent among themselves at both the primary and secondary structure levels as predicted by hydrophilicity profiles (Hopps and Woods algorithm)³. Table 1 shows the alignment of the amino-acid sequences, while Fig. 1 shows that the relative position of the regions of similar sequence in the respective proteins is the same. The similarities in sequence and sequence organization may reflect either a common evolutionary origin or common function. It is noteworthy that the only region of sequence similarity between EPA and the human granulocyte-macrophage colony-stimulating factor⁴, besides a short hydrophobic stretch close to the amino terminus, is region III, which is also the most highly conserved region among the retroviral core proteins.

The similarities reported here open the possibility that the major gag core protein may be important not only in viral assembly but also in affecting the cellular physiology so as to favour viral replication



Fig. 1 Regions of similarity between EPA-TIMP and the major retroviral gag core proteins. aa, Amino acids.

acting as a growth factor or a protease inhibitor. In the latter case it could prevent the breakdown of proteins essential for viral replication such as reverse transcriptase or the endonuclease-integrase. These proteins are exposed to the cytoplasmic and nuclear environments during the early stages of the infection process. The major retroviral core protein may also regulate changes in the cytoskeleton during the process of budding. In this regard we note that certain pathogenic determinants of some avian and murine retroviruses have been mapped to the gag-pol region^{5,6}. Using an analogy with the gag core protein function, EPA-TIMP may interact with RNA and with membrane proteins in its regulatory activity.

Many of the transforming oncogenes transduced by retroviruses are synthesized as fusion proteins that contain a substantial portion of the major capsid protein. The deletion of the gag determinants in some cases results in loss of transforming activity¹³ which could be due to loss of growth-promoting or protease inhibitory activities associated with the gag portion of the fusion product.

ROBERTO PATARCA
WILLIAM A. HASELTINE

Laboratory of Biochemical Pharmacology,
Dana-Farber Cancer Institute,
Department of Pathology,
Harvard Medical School,
Boston, Massachusetts 02115, USA

1. Gasson, J. C. *et al.* *Nature* **315**, 768-771 (1985).
2. Docherty, A. J. P. *et al.* *Nature* **318**, 66-69 (1985).
3. Hopps, T. P. & Woods, K. R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3824 (1981).
4. Lee, F. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **82**, 4360-4364 (1985).
5. Robinson, H. L., Blais, B. M., Tschlis, P. W. & Coffin, J. M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1225-1229 (1982).
6. Holland, C. A., Hartley, J. W., Rowe, W. P. & Hopkins, N. J. *J. Virol.* **53**, 158-165 (1985).
7. Ratner, L. *et al.* *Nature* **313**, 277-284 (1985).
8. Seiki, M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **80**, 3618-3622 (1983).
9. Shimotohno, K. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **82**, 3101-3105 (1985).
10. Sagata, N. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **82**, 677-681 (1985).
11. Shinnick, T. M., Lerner, R. A. & Sutcliffe, A. *Nature* **293**, 543-548 (1981).
12. Shwarts, E., Tizard, R. & Gilbert, W. *Cell* **32**, 853-869 (1983).
13. Prywes, R. *et al.* *Cell* **34**, 569-579 (1983).

Table 1 Alignment of amino-acid sequences in regions of similarity shown in Fig. 1

Region	Origin	Position	Sequence
I	HTLV-III	35	E V I P M F S A L S E G A T P
	HTLV-I	32	P G S P Q F M Q T I R L A V Q
	HTLV-II	32	E G S P Q F M Q T I R L A V Q
	BLV	32	P G S Q V W I Q T L R L A I L
	Mo-MuLV	34	V E P G K L T A L I E S V L I
	RSV	33	I T M A E V E A L M S S P L L
	EPA	-2	P T M A P F E P L A S G I L L
II	HTLV-III	112	Q E Q I G W M . T N N P F I P V G E I
	HTLV-I	112	E Y Q Q L W L . A A F A A L P G S A K
	HTLV-II	112	E T Q N L W L . A A F S T L P G N T R
	BLV	112	Q Y Q N L W L Q A G K I S L L V L Q L
	Mo-MuLV	109	Q L L L A G L Q N A G R S P T N L A K
	RSV	116	Q G Q A A L L R P G E L V A I T A S A
	EPA	75	L G D A A D I R F V Y T F A M E S V C
III	HTLV-III	150	I L D I R Q G P K E F F R D Y V D R F
	HTLV-I	133	W A S I L Q G L E E P Y H A F V E R L
	HTLV-II	133	W A A I L Q G L E E P Y C A F V E R L
	BLV	133	W S T I V Q G P A E S S V E F V N R L
	Mo-MuLV	130	ITQGPN . . I T Q G P N E S P S A F L E R L
	RSV	149	W A D I M Q G P S E S F V D E A N R L
	EPA	113	Q D G L L H I T T C S F V A P W N S L S
IV'	HTLV-III	193	N A N P D C K T I L K A L G
	HTLV-I	185	N A N K E C Q K I L Q A R G
	HTLV-II	185	N A N K E C Q K L L Q A R G
IV	RSV	191	F R Q K S Q P D I Q Q
	EPA	153	F P C L S I P C K L Q

Similarities among the major gag core proteins of HTLV-III⁷, HTLV-I⁸, HTLV-II⁹, BLV¹⁰, Mo-MuLV¹¹ and RSV¹² and EPA¹ or TIMP² and the granulocyte-macrophage colony-stimulating factor⁴ are shown. Common amino acids and conservative substitutions are boxed. A repeat is present in region III in Mo-MuLV.

Body size and biomass

A RECENT article¹ described several examples of the double logarithmic relationship between the number of individuals and their body length for mixed arthropod populations resident on vegetation of various surface fractal dimensions. Eight examples were given graphically spread over five diagrams. The mean of the slopes of the regression lines through the diagrams can be derived, from the individual slopes given¹, as -3.028 ; that is within 1% of -3 . A slope of -3 indicates a relation of the form: $\log N = \log K - 3 \log L$, where N is the number of individuals, K is constant and L is a body length; that is, $N = KL^{-3}$.

However, if we assume a constant body shape and that all arthropods enjoy a uniform density for a first approximation, we have: $m = k_0 L^3$, where m is body weight and k_0 is a constant. This converts to $L = k_1 m^{1/3}$, where k_1 is a constant. Therefore, $N = K(k_1 m^{1/3})^{-3} = k_2 m^{-1}$ where k_2 is a constant. But $m = M/N$, where M is the total biomass of all N individuals of a given size. Therefore, $N = k_2(M/N)^{-1} = k_2(N/M)$; that is, $M = k_2$. In other words, the biomass of individuals of one size, summed over all species within a community of mixed arthropods, is constant.

I thank Professor J. H. Lawton for helpful discussions on this matter.

BRUCE DENNESS

Bureau of Applied Sciences,
Wydcombe Manor,
Whitwell,
Isle of Wight PO38 2NY, UK

I. Morse, D. R., Lawton, J. H., Dodson, M. M. & Williamson, M. H. *Nature* 314, 731-733 (1985).

Polarity reversal in the Solomon Islands arc

IN a recent letter, Cooper and Taylor¹ present evidence for two juxtaposed Wadati-Benioff zones in the Solomon Islands, and state that the seismic data provide the first direct evidence of a reversal of subduction polarity at an island arc.

The concept of arc polarity has received considerable attention in the literature²⁻⁴, particularly as such studies suggest how the tectonic settings of former volcanic belts might be interpreted. Cooper and Taylor, following the observations of other researchers on the two seismic zones in the Solomons⁵⁻⁸ raise one of the fundamental tectonic problems for this region, but in doing so fail to give perspective to their proposal by highly selective referencing—ignoring many contributions which might negate or modify their conclusions—and by greatly simplifying the dating of Tertiary volcanism in the region.

Central to the thesis of Cooper and Taylor is the concept that the Ontong Java Plateau collided with the North Solomon Trench sometime during the Miocene. A number of researchers who have been actively working in the Solomons have, over the last decade or so, questioned to varying degrees the concept of collision⁹⁻¹³ and there are strong geological grounds for suggesting that the Solomon Islands may represent the leading edge of the Ontong Java Plateau¹³⁻¹⁵. If two initially quite separate crustal blocks collided, the question rises as to where this collision zone is to be located so that it is in accordance with the known geology.

Arc polarity reversal in the Solomons is often associated with the collision model, following a widely quoted paper by Halunen and Von Herzen¹⁶. Yet what has received far less attention is that one of the major pieces of evidence cited by Halunen and Von Herzen in support of arc polarity reversal (high heat flow on the south-west flank of the Solomons) is more likely associated with the Woodlark Spreading Zone^{17,18}. The presence of an incipient seismic zone dipping south-west below Santa Isabel is interesting, yet its absence to the north along Bougainville requires special pleading by Cooper and Taylor. That this incipient seismic zone is a remnant subduction zone is one of a number of possibilities and until new geological evidence is forthcoming, I agree with Coleman¹⁹ that the concept of north-east-facing arc polarity in the Solomons during the Palaeogene is not proven.

Cooper and Taylor state, as support for their model of collision and polarity reversal, that there was a hiatus in arc volcanism lasting from late-early to early-late Miocene. However, on many of the islands, evidence for the age of volcanism is very poor and no clear age groupings may be given. On Bougainville the Kieta volcanics are dated as Oligocene? to lower Miocene²⁰, though it is noted that unconsolidated equivalents at Kanga Hill range through to possible Pliocene age. Unnamed volcanics are equated in age with either the older Kieta volcanics or the younger Emperor Range volcanic beds merely on their location and erosional state, while the Buka Formation is dated on geomorphic evidence alone.

In the Shortland Group, volcanism is considered to be pre-Pliocene²¹; Turner and Ridgeway emphasize the lack of fossil and radiometric age control on the various volcanic episodes. It has been suggested that in the New Georgia Group volcanism commenced during the Pliocene¹⁸, though there is evidence for volcanism extending well back into the Miocene^{21,22}. On Choiseul²³ the oldest preserved post-base-ment sediments were deposited in the late Oligocene to early Miocene and, during early and middle Miocene, these contained increasing amounts of andesitic volcanic material. Possibly the best

evidence for a hiatus in arc volcanism comes from Guadalcanal²⁴ where the top of the Suta volcanics is placed on the Aquitanian-Burdigalian boundary. However, the age of the base of the Gold Ridge volcanics is uncertain and Hackman²⁴ notes that Suta-type volcanism continued intermittently in north-central Guadalcanal into the Pliocene. For San Cristobal, any deduction regarding the tectonic position of the island from Oligocene time onward that is based solely on the absence of arc volcanism in its stratigraphical record, is likely to be incorrect. In complex island arc settings, such as the Solomons, with rapid facies variation, localized unconformities, spasmodic volcanic episodes, and the paucity of age controls on volcanism, it is premature to make emphatic statements about breaks in volcanism and then to use the assertions to lend support to the notion of collision and arc polarity reversal.

The suggestion by Cooper and Taylor that volcanism on Savo may be the product of reactivated magmatism associated with a relic slab is not supported by petrological data¹³. The broad similarities in petrography, mineralogy and geochemistry between lavas from Savo, north Guadalcanal and Vella Lavella suggest, as a working hypothesis, that the same tectonic and magmatic phenomena have been responsible for the style of volcanism in these three localities.

Finally, I draw attention to the warnings of Arculus and Johnson²⁵ and make a plea that tectonic models proposed for island arcs such as the Solomons should balance generalized arc models carefully with the unique features as displayed in the field evidence and field relations of this enigmatic fractured island chain.

W. R. H. RAMSAY

State Geological Survey of Victoria,
PO Box 173, Melbourne,
Victoria 3002, Australia

- Cooper, P. & Taylor, B. *Nature* 314, 428 (1985).
- Kuno, H. *Bull. volcan.* 20, 37-76 (1959).
- Jakes, P. & White, A. J. R. *Geochim. cosmochim. Acta* 34, 849-856 (1970).
- Gill, J. & Gorton, M. in *The Western Pacific: Island Arcs, Marginal Seas, Geochemistry* (ed. Coleman, P. J.) 543-566 (University of Western Australia Press, 1973).
- Westwood, J. V. B. *J. geol. Soc. Aust.* 17, 87-92 (1971).
- Curtis, J. W. *J. geol. Soc. Aust.* 20, 1-19 (1973).
- Denham, D. *Bull. Aust. Soc. Explor. Geophysicists* 6, 78-79 (1975).
- Neef, G. *Bull. Aust. Soc. Explor. Geophysicists* 9, 99-103 (1978).
- Hackman, B. D. in *The Western Pacific: Island Arcs, Marginal Seas, Geochemistry* (ed. Coleman, P. J.) 179-191 (University of Western Australia Press, 1973).
- Coleman, P. J. & Hackman, B. D. *Geol. Soc. Lond. Spec. Publ.* No. 4, 453-461 (1974).
- Ramsay, W. R. H. *Bull. Aust. Soc. Explor. Geophysicists* 9, 107-110 (1978).
- Turner, C. C. & Hughes, G. W. *Tectonophysics* 87, 127-146 (1982).
- Ramsay, W. R. H. thesis, Univ. New England (1981).
- Ramsay, W. R. H. *Tectonophysics* 87, 109-126 (1982).
- Ramsay, W. R. H. *15th Pacific Sci. Congr. Abstr.* 2, 193 (1983).
- Halunen, A. J. & Von Herzen, R. P. *J. geophys. Res.* 78, 5195-5208 (1973).
- Milsom, J. S. *J. geophys. Res.* 75, 7335-7339 (1970).

18. Ramsay, W. R. H., Crawford, A. J. & Foden, J. D. *Contr. Miner. Petrol.* **88**, 386-402 (1984).
19. Coleman, P. J. *Bull. Aust. Soc. Explor. Geophysicists* **6**, 60-61 (1975).
20. Blake, D. H. & Miezitis, Y. *Bull. Bur. Miner. Resources Geol. Geophys.* **93**, (1967).
21. Turner, C. C. & Ridgeway, J. *Tectonophysics* **87**, 335-354 (1982).
22. Stanton, R. L. *Overseas Geol. Miner. Resour.* **8**, 127-149 (1961).
23. Ridgeway, J. *Br. Tech. Co-op., West Sol. Isl. Geol. Mapping Project* **19** (1984).
24. Hackman, B. D. *Inst. Geol. Sci. Overseas Mem.* **6**, (1980).
25. Arculus, R. J. & Johnson, R. W. *Earth planet. Sci. Lett.* **39**, 118-126 (1978).

TAYLOR AND COOPER REPLY—Ramsay rightfully asserts that we did not cite many conflicting articles on the geology of the Solomon Islands in the short introduction to our paper¹.

There are two completely opposite interpretations of Solomon Islands geology. Ramsay² and Turner³ champion the interpretation that the Cretaceous oceanic basement on Malaita, Santa Isabel, Choiseul and Guadalcanal is similar, that the post-Eocene sediments on Malaita and Santa Isabel include fine-grained products of volcanism on Choiseul and Guadalcanal, and that an Oligocene arc was built on the leading edge of the Ontong Java Plateau above a north-east-dipping subduction zone. In contrast, Kroenke^{4,5} and Coleman^{6,7} champion the interpretation that whereas the unmetamorphosed basement and pelagic sediments of Malaita and Santa Isabel are similar to those of the Ontong Java Plateau, they are quite different from the metamorphosed basement and volcanoclastic sediments of Choiseul and Guadalcanal, requiring a Miocene collision between the two provinces, involving obduction of the former above a south-west-dipping subduction zone. The first group interpret the *en echelon* fault-bounded, peridotite-gabbro Korigohle trend as a structural high forming a sediment barrier between the two provinces, whereas the second group interpret the same feature as the suture between the two provinces. The geology is sufficiently complicated that some authors including Coleman^{6,8} and Hughes and Turner^{3,9} have published self-contradictory interpretations. Thus, the evidence from Solomon Islands geology alone for a north-east-facing middle Tertiary arc which collided with the Ontong Java Plateau in the Miocene is ambiguous⁸ and open to conflicting interpretations^{2,3,5,7}.

Other evidence presented by Halunen and Von Herzen¹⁰ in support of arc polarity reversal includes the high heat flow and high shear-wave attenuation on the Solomon Sea side of the arc, the location of active volcanoes on the south-west side of the islands very close to the New Britain-San Cristobal trench, and the presence of deep (>400 km) earthquakes beneath the islands. Taylor and others^{8,11,12} have already noted that most of these phenomena are the result of active seafloor spreading in the Woodlark Basin and subduction of the spreading ridge

beneath the island arc. The deep earthquakes can be related to present subduction beneath the north-west Solomons and to either present north-east or past south-west subduction beneath the south-east Solomons¹³.

Ramsay also points out the lack of radiometric age dates on igneous rocks in the Solomon Islands and the fact that most of the volcanic formations are dated by stratigraphic and fossil evidence, often with poor control. Nevertheless, the early and middle Miocene is a rather well-constrained period in which the widespread deposition of limestones, calcarenites and foraminiferous marls (the Hautara, Mbetilonga and Mbonehe, and Keriatu limestones on San Cristobal, Guadalcanal and Bougainville, respectively) mark a regional gap in the volcanic record^{5,6}. There are locally intercalated volcanoclastic sequences, and extensive greywackes on Choiseul and Guadalcanal, but these appear to be derived from erosion of earlier volcanics^{5,6,14}. There are no dates older than 7 Myr on igneous rocks associated with the present north-east subduction⁵. Although we noted that the apparent middle Miocene hiatus in volcanism may be evidence of a temporary cessation of subduction during the polarity reversal¹, we did not emphasize this point because the evidence from the Philippines and other areas indicates that volcanism and even double-sided subduction may occur during an arc polarity reversal.

What then is the remaining evidence for arc polarity reversal in the Solomon Islands? The term 'arc polarity reversal' was coined by Karig and Mammerickx¹⁵, who were the first to propose that this process had occurred in the Solomon Islands, primarily on the basis of studies of the New Hebrides along-strike. They assumed that all marginal basins were formed by crustal extension behind trenches, and therefore that the New Hebrides and Solomon Islands arcs formerly faced north-east. Although this assumption has since been proven incorrect^{16,17}, it is still true that the geology and geophysics of the island arcs along-strike from the central Solomons provide the least controversial evidence for arc polarity reversal. Petrochemical, sedimentological, palaeomagnetic and radiometric age data from the New Hebrides and Fiji all confirm a north-east-facing Vitiaz arc in the late Oligocene-early Miocene¹⁸⁻²¹. Vitiaz arc volcanism ceased in the late-early Miocene, the area was rifted in the middle Miocene, and following arc polarity reversal at ~10 Myr BP, a new arc was built, beginning at 7 Myr BP¹⁸⁻²¹. The Vitiaz trench and adjacent north-east edge of the Fiji Plateau still preserve the classical trench/fore-arc morphology/structure of the former north-east-facing arc²²; the same is true of the Kilinailau-Manus trench and the remnant forearcs north and east of Bougainville, New Ireland and Manus, on the other side of the

central Solomon Islands^{23,24}. Therefore, although some arguments used by early authors^{10,15} in favour of arc reversal have since been disproven, reversal of the island arcs adjacent to the central Solomons remains well documented by both island geology and marine geophysical evidence. An independent mid-Tertiary evolution of the central Solomons on the Pacific Plate would require over 1,000 km of differential movement with respect to the surrounding arcs with which they have long been associated⁶.

Given this context, we reaffirm our previous conclusion that the two juxtaposed Wadati-Benioff zones of opposite polarity provide direct seismic evidence of a reversal in subduction polarity at the Solomons island arc¹. Furthermore, although intermediate-depth seismicity on the south-dipping remnant subduction zone is only observed adjacent to the young Woodlark Basin, significant shallow seismicity is associated with this feature to the north-west^{1,25}. Finally, the location of recent volcanism on Savo, 100-150 km above the relic slab, surely implies some genetic relation between the two.

B. TAYLOR
P. COOPER

*Hawaii Institute of Geophysics,
University of Hawaii at Manoa,
2525 Correa Road, Honolulu,
Hawaii 96822, USA*

1. Cooper, P. & Taylor, B. *Nature* **314**, 428-430 (1985).
2. Ramsay, W. R. H. *Tectonophysics* **87**, 109-126 (1982).
3. Turner, C. C. & Hughes, G. W. *Tectonophysics* **87**, 127-146 (1982).
4. Kroenke, L. W. thesis, Univ. Hawaii (1972).
5. Kroenke, L. W. *Cenozoic Tectonic Development of the South-west Pacific* (CCOP/SOPAC Tech. Bull. **6**, 1985).
6. Coleman, P. J. & Packham, G. H. *Earth Sci. Res.* **12**, 197-233 (1976).
7. Coleman, P. J. & Kroenke, L. W. *Geo-Marine Lett.* **1**, 129-134 (1981).
8. Coleman, P. J. *Pap. presented at IDOE Workshop Suva, Fiji, 1-6 September 1975* (CCOP/SOPAC Tech. Bull. **2**, 1975).
9. Hughes, G. W. & Turner, C. C. *Bull. geol. Soc. Am.* **88**, 412-424 (1977).
10. Halunen, A. J. & Von Herzen, R. P. *J. geophys. Res.* **78**, 5195-5208 (1973).
11. Weissel, J. K., Taylor, B. & Karner, G. D. *Tectonophysics* **87**, 253-277 (1982).
12. Taylor, B. in *Earth Science Ser.* (eds Taylor, B. & Exon, N.) (Circum. Pacific Council for Energy and Mineral Resources, in the press).
13. Cooper, P. thesis, Univ. Hawaii (1985).
14. Hackman, B. D. *Inst. Geol. Sci. Overseas Mem.* **6**, (1980).
15. Karig, D. E. & Mammerickx, J. *Mar. Geol.* **12**, 187-205 (1972).
16. Weissel, J. K. *Phil. Trans. R. Soc. A300*, 223-247 (1981).
17. Taylor, B. & Karner, G. D. *Rev. Geophys. Space Phys.* **21**, 1727-1741 (1983).
18. Gill, J. B. & Gorton, M. P. in *The Western Pacific: Island Arcs, Marginal Seas, Geochemistry* (ed. Coleman, P. J.) 543-566 (University of Western Australia Press, 1973).
19. Falvey, D. *Bull. Aust. Soc. Explor. Geophysicists* **9**, 117-123 (1978).
20. Carney, J. N. & Macfarlane, A. *Bull. Aust. Soc. Explor. Geophysicists* **9**, 123-130 (1978).
21. Carney, J. N. & Macfarlane, A. *Tectonophysics* **87**, 147-175 (1982).
22. Brocher, T. M. in *Earth Science Ser.* **3** (ed. Brocher, T. M.) (Circum. Pacific Council for Energy and Mineral Resources, in the press).
23. Kogan, M. G. *J. geophys. Res.* **81**, 5240-5248 (1976).
24. de Broin, C. E., Aubertin, F. & Ravanne, C. in *Int. Symp. Geodynamics of the Southwest Pacific New Caledonia, August-September 1976*, 33-47 (Editions Technip, 1977).
25. Weissel, J. K. & Anderson, R. N. *Earth planet. Sci. Lett.* **41**, 143-158 (1978).

The immunology industry

The wide range of monoclonal antibodies now available includes murine antibody to human calcitonin and antibodies to cytokeratin proteins.

• Becton Dickinson's monoclonal antibody Anti-Leu 12 (clone 497) is a B-cell marker for researchers studying B cells in blood or tissue section. This antibody recognizes a unique antigen present on all B cells as defined by the presence of kappa or lambda light chains. Anti-Leu 12 has not been found to react with normal T cells or concanavalin A-stimulated T cells, or granulocytes and has only marginal reactivity with monocytes and tissue macrophage. The reagent is available in a purified form or with a fluorescein conjugate and is supplied in a liquid ready-to-use solution with a long life at 4°C.

Reader Service No. 100.

• Preparative FPLC 100 and 300 ml Monobead columns are now available from Pharmacia. The IFPLC-10µ system uses new P-6000 pumps for increased high performance throughout. Designed for the separation of biological molecules, FPLC columns are widely used for the purification of monoclonal antibodies, products from recombinant DNA technology and other biologicals.

Reader Service No. 101.



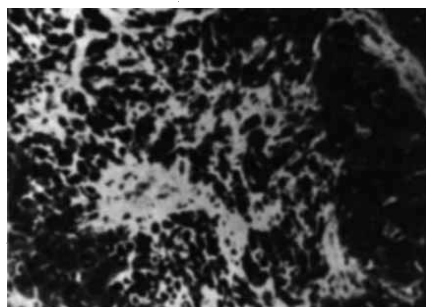
Pharmacia's chromatography system, suitable for purification of monoclonals.

• A new series of Vectastain avidin-biotin complex (ABC) kits, with alkaline phosphatase (AP) as the marker, are now available in the UK for immunohistochemical staining tests. Based on the principal of the pre-formed avidin/biotinylated enzyme complex, the system is recommended in situations where high sensitivity is a prerequisite or when a high level of endogenous peroxidase activity precludes the use of the peroxidase-based ABC system. These kits can also be employed in double labelling techniques with other Vectastain ABC systems — peroxidase or glucose oxidase — and for enzyme immunoassays and nitrocellulose techniques. Following the same format as other Vectastain kits, the ABC-AP kit contains blocking serum, biotinylated secondary

These notes are based on information provided by the manufacturers. To obtain further details about these products use the reader service card bound inside the journal.

antibody, avidin DH and biotinylated alkaline phosphatase H. For convenience, a standard ABC-AP kit (with only avidin DH and biotinylated alkaline phosphatase H) is also available. Adjunctive reagents include three substrate kits, which offer a choice of three coloured precipitates: blue, black or Vector Red — a bright red permanently mountable colour.

Reader Service No. 102.



The Vector alkaline phosphatase staining system.

• The Clontech Laboratories line of cDNA libraries has been expanded to include lambda gt 11 cDNA libraries from the following tissues: human brain, human fetal liver, human lung, human neuroblastoma, bovine adrenal, bovine retina and rat lens. A genomic library from human leukocyte is also available in EMBL-3. Custom cDNA library construction is also available.

Reader Service No. 103.

• The Litton Bionetics range includes monoclonal antibodies directed against murine major histocompatibility complex antigens (H-2, Ia) from 26 cell lines. These antibodies recognize H-2 and Ia specificities coded for by genes of several haplotypes. The product is of three types — supernatant, clarified ascites or purified immunoglobulins — and serological specificity is assured by performing a complement-mediated, dye-exclusion cytotoxic assay of each product, using a panel of eight strains of mice representing non-recombinant and recombinant H-2 haplotypes. Rabbit serum screened for low toxicity to mouse splenocytes is used as the source of complement. Litton's products are available through Universal Biologicals in the UK.

Reader Service No. 104.

• Kirkegaard & Perry Laboratories now produce enzyme immunoassay kits for the detection of mouse monoclonal antibodies. The Hybri-Clonal EIA screens hybridoma cultures for clones that produce antibodies of desired specificity. The kit contains peroxidase labelled affinity purified antibody and all the reagents for 400 tests. Hybri-Clonal kits enable the

selection of clones producing antibodies of either IgM or IgG class and/or both simultaneously.

Reader Service No. 105.

• LKB Instruments has produced a 20-page technical description of their 1.2 and 4 detector automatic gamma counter, the model 1272 CliniGamma. Cutaway drawings show how the shielded detectors in this multiple-detector RIA counter are positioned. Other sections deal with the continuous automatic calibration of all detectors, standardization of detector efficiencies, backgrounds and inter-detector crosstalk, plus chi square testing for detector repeatability. Sample racks, containers, variation in sample volumes, compatibility with automatic pipettors and centrifuges, and a positive sample ID system are all detailed. Programming and control of the CliniGamma receives over five pages of explanation. Data reduction methods given are: log-logit regression and the sophisticated cubic spline function for curve fitting plus various emulations of manual point-to-point graphing, a thyroid group for T₄/FTI/free T₄, and a hepatitis screening group of five methods, all on- or off-line.

Reader Service No. 106.



Tissue culture and monoclonal production are applications requiring pure water, as produced by the Millipore Milli-Q.

• The Millipore Milli-Q Water System provides organic-free (less than 20 parts per 10⁶ TOC) or pyrogen-free 18-megohm water for critical applications including tissue culture and monoclonal antibody production. The Milli-Q system produces just as much water as needed, when needed, at up to 1.5 litres a minute. This eliminates the need for storage and the consequent degrading of water quality. Unlike stills which require regular time-consuming scale removal, Milli-Q cartridges can be changed in minutes without tools. In addition, Milli-Q Ultrapure cartridge kits provide all the replacement cartridges needed under one catalogue number.

Reader Service No. 107.

ADVERTISEMENTS

LIPID PRODUCTS

top quality lipids for biomedical research
17 years experience supplying lipids to top
research laboratories throughout the world

For free catalogue contact:
Lipid Products, Nutfield Nurseries
South Nutfield, Redhill
Surrey RH1 5PG, UK
Tel: (073 782) 3277

Reader Service No.19

LIPOSOMES

REPRODUCIBLE - HOMOGENEOUS -
EXTREMELY RAPID

Our new commercial LIPOSMAT
guarantees rapid (5ml within 60 minutes)
preparation of uniformly sized

- UNILAMELLAR OR
- OLIGOLAMELLAR OR EVEN
- MULTILAMELLAR LIPOSOMES

depending on the used experimental
conditions. Liposome size is selectable
between 25 and ca. 600 nm diameter up
to 1000 nm for multilamellar liposomes.
Sample volume is normally 5-10ml or up
to 200ml if used in combination with an
additional instrument. High encapsulation
efficiency for hydrophilic and/or lipophilic
agents. Lipid concentration can be up to
300 mg/ml. Lipid loss during preparation
is extremely low: down to 0.1%.

BIANORM - Geräte

POB 126, D-8000 Munich 65, FRG

Reader Service No.36

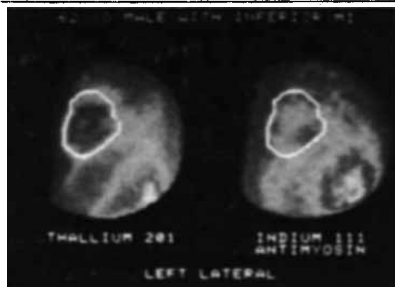
Record The

LUNAR COMPONENT

In your research, with The Lunatime —
The only 24 hr. 50½ min quartz clock
made. (It's Circadian!)

Money back guarantee by the makers of
the Tide Timer, the world's largest sell-
ing Tide Clock. Immediate shipment.
Send \$50. U.S. + \$2. UPS (or \$10. over-
seas air.) Schelling Research Corp. P.O.
Box 177 Newton, N.J. 07860 (201) 383-
8467

Reader Service No.43



Cardiac diagnosis by Centocor.

• Centocor Inc. has introduced a
radiolabelled monoclonal antibody for
rapid differentiation of cardiac necrosis
and ischaemia following myocardial in-
farction. Called Myoscint, the antibody is
selective for the heavy chain of the in-
tracellular cardiac muscle component
myosin which is exposed when the cell
membrane ruptures during a heart attack.
The indium-111 labelled Myoscint there-
fore binds exclusively to necrotic cardiac
cells and images of these are obtained us-
ing a conventional gamma camera.

Reader Service No. 108.

• Serotec of Bicester, Oxfordshire, has
announced an extended range of con-
jugated monoclonal antibodies. The com-
pany developed a method of conjugation
which makes it possible to add labels to
most antibodies in the Serotec range of
monoclonal products. These labels in-
clude FITC, TRITC, peroxidase and alka-
line phosphatase. The conjugates are sup-
plied in 1 ml vials, with a concentration of
bound antibody of 1 mg ml⁻¹.

Reader Service No. 109.

• Hybritech's list of monoclonal anti-
bodies has been expanded to include
murine monoclonal antibody to human
calcitonin, used in studies of certain carci-
nomas and bone disease, and monoclonal
antibody to bombesin, a tetradeca-
peptide, present in mammalian tissues
including brain and gut. Also available
now is pooled AE1 and AE3 monoclonal
antibodies to human epithelial keratin,
which should prove useful for the positive
identification of normal and abnormal
epithelial cells in tissue sections and cul-
ture, and for distinguishing carcinomas
from non-carcinomas. And monoclonal
antibodies to hepatitis B surface antigen
and subtypes as, adw and ay are also avail-
able, which should help in a new and im-
proved methods for hepatitis screening
and diagnosis.

Reader Service No. 110.

• A new HistScan monoclonal detector
kit for the localization of both mouse and
rat monoclonal antibodies following their
reaction with tissue antigens is now avail-
able from Biomed. This monoclonal de-
tector kit is designed to simplify the tissue
immunostaining technique using mono-
clonal antibodies specific for tissue-
associated antigens. It is based on the new
avidin biotin complex technology kits con-
tain all reagents needed to detect mono-
clonal antibodies in 150 tissue sections.
Also contained in each kit are haematoxy-
lin counter-staining reagent, concentrated
chromogen reagents and pertinent block-
ing solutions to minimize nonspecific
staining. The reagents forming the kit are
supplied in a ready-to-use form together
with a step-by-step procedure outline and
other technical information. When stored
at 4°C, the kit has a shelf life of one year.

Reader Service No. 111.

• A new edition of the Boehringer Man-
nheim Biochemicals Catalogue is now
available. It lists over 1,500 products in-
cluding products for molecular biology,
immunology, cell culture and general
biochemistry. This revised edition is de-
signed to be easy to use. The main section
is an alphabetical product listing which
contains all the pertinent information on
product specifications. In addition, there
are quick reference listings for molecular
products and for immunology and cell-
culture products, extensive cross refer-
ences and indexes organized both
alphabetically and by product group. For
a free copy contact Boehringer Mannheim
at 7941 Castleway Drive, P.O. Box 50816,

Indianapolis, Indiana 46250.

Reader Service No. 112.



Tago's new reagents for immunohistology.

• Tago's series of horseradish
peroxidase-antiperoxidase (PAP) soluble
complexes has been specially prepared for
use in immunohistological staining and
quantitative enzyme immunoassays.
These convenient, step-saving reagents
provide a more sensitive assay than
fluorochrome-conjugated antibodies and
a higher signal than standard peroxidase-
conjugated antibody. They also may be
used with fixed tissues as well as with
frozen sections in detecting most antigens.
PAP complexes are now available for ap-
plications with goat or rabbit primary im-
munoglobulins. They are effective at
working dilutions of 1:20,000 for ELISA
and up to 1:200 for tissue staining proce-
dures. PAP complexes can be linked with
primary antibodies through second anti-
bodies to goat or rabbit IgG.

Reader Service No. 113.

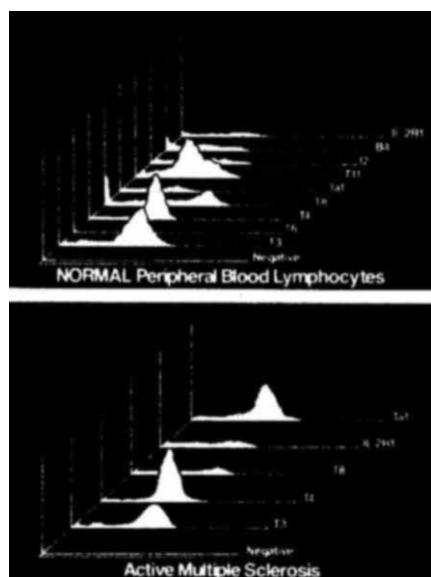
• Dako Corporation is introducing con-
trol slides in 10-packs — one stained for
reference and nine unstained slides. These
ready-prepared controls can reduce time
spent searching for, acquiring and process-
ing tissue. Hard-to-obtain controls for
antigens such as α -fetoprotein, papillo-
mavirus, herpes simplex virus and nearly
20 other immunohistochemical markers
are included in this new product line. Each
slide is ready for immediate use. Each kit
includes recommended step-by-step proce-
dure and a technical staining description.

Reader Service No. 114.

• Adding to its range of arachidonic acid
metabolites, Seragen has introduced a
leukotriene B₄ ³H RIA kit. With its highly
specific sensitivity, the researcher can
achieve accurate results with samples as
low as 10 pg per 100 μ l in a simple 2-hour
room temperature assay. The antiserum
has low cross-reaction with other leuko-
triens and HETEs so that no HPLC sam-
ple separation is necessary.

Reader Service No. 115.

• Coulter Immunology has announced
the availability of three exclusive leuko-
cyte monoclonal antibodies. The new
Coulter Clone monoclonal antibodies —
2H4-RD1, 4B4-RD1 and Ta1-RD1 — will
enable investigators to identify functional
T4 subsets which play precise roles in the
onset of autoimmune and immune de-
ficiency diseases. Anti-2H4-RD1 defines
the functionally unique T4+ subset, sup-
pressor inducer, which falls dramatically
in certain patients with the onset of syste-
mic lupus erythematosus. 4B4-RD1 de-
fines the T4+ cell subset, helper inducer,
and may be the key parameter for the



Results with Coulter's new antibodies.

evaluation of immune deficiencies such as AIDS. These major subsets can be identified by simultaneous analysis with Coulter Clone T4-FITC. T4 is a cell-specific activation antibody available only from Coulter. A recent study of multiple sclerosis patients demonstrated positive correlation between increased T4+ cells in the peripheral blood and systemic immune activation in the disease.

Reader Service No. 116.

- A new Pentex antigen/antibody enhancement medium and antisera diluent called Mod-U-Cyte is now available for research and manufacturing use at two protein concentrations, 6% and 20%. The suppliers, Miles Scientific, say its advantages include: Rh and certain other rare antibodies are more reactive at both pre- and post-antiglobulin test phases when compared to conventional, highly polymerized and low ionic strength bovine albumin solutions; certain weak clinically significant IgM antibodies can be detected without using proteolytic enzyme-premodified red blood cells; and with Mod-U-Cyte as a blood grouping and typing antisera diluent, most titres are extended two to 16 times when compared to solutions using bovine albumin as a diluent. Mod-U-Cyte may enhance availability of cell surface antigen sites.

Reader Service No. 117.

- Jackson ImmunoResearch has introduced a new line of over 200 biotinylated antibodies which have a 13-atom spacer between biotin and the antibody. This results in a significant increase in sensitivity. The spacer appears to extend the biotin moiety far enough from the antibody surface to make it more accessible to biotin-binding sites on streptavidin and avidin, especially when they are labelled with high molecular weight enzymes. It appears that either more enzyme-avidin conjugate can be spatially accommodated on the antibody surface or perhaps the extended biotin has better access to its

binding sites which are sterically recessed from the surface of the enzyme-covered avidin. The new products, designated Biotin-SP to indicate biotin with a spacer, will be available on all biotinylated AffiniPure antibodies and ChromPure proteins. Biotin-SP enzymes containing the same 13-atom spacer will also be available to form avidin-biotin-enzyme complexes for use in histochemical/cytochemical procedures requiring ultra-high sensitivity. Fluorophore-labelled and enzyme-labelled streptavidin and avidin are available for use with Biotin-SP-AffiniPure antibodies and Biotin-SP-ChromPure proteins.

Reader Service No. 118.

- Celltech Limited produces a range of immunoprocessing reagents based on the company's expertise in large scale monoclonal antibody production. In September 1985 Celltech extended its range of monoclonal-antibody-based immunoprocessing reagents with the introduction of Resolute IL-2 and IL-2 IRMA, for the purification and assay, respectively, of human interleukin-2. The reagent gives a high purity of interleukin-2 in a single-step chromatographic process, and is a reliable, rapid and reproducible alternative to the more traditional 'bioassay' in-2. Celltech also supplies monoclonals for ABO blood typing.

Reader Service No. 119.

- Euroclone is a new European company specializing in the contract production of mammalian secreted products such as monoclonal antibodies, human growth hormone and tissue plasminogen activator. The company is interested in both small and large scale production orders. Production is based on the computerized facility, Acusyst-P. This system, manufactured by Endotronics Inc., uses hollow-fibre technology in combination with a dosage of cell media. For each customer all the manufacturing conditions are very well defined and reproducible. The cost per gram of secreted product should be significantly lower than the usual level. Euroclone is a joint venture of Dalton-Waalwijk, of the Netherlands, and a well-known European immunology institute.

Reader Service No. 120.

- BioTope, Inc. of Bellevue, Washington has announced two new immunoassay kits for the cancer research market. These assay kits are claimed to be the first commercially available immunoassays for detecting and measuring human transforming growth factor alpha (hTGF- α) and human epidermal growth factor (hEGF). The kits detect these growth factors in urine. Based on BioTope's proprietary microassay technology, BioMAT, the kits combine speed, economy and reproducibility. BioMAT should find numerous applications in immunoassays, receptor assays and DNA probe assays.

Reader Service No. 121.

- A radioimmunoassay kit suitable for the determination of the level of atrial

natriuretic polypeptide (ANP) in unextracted plasma samples has been developed by Research & Diagnostic Antibodies. ANP is the recently described heart hormone involved in the regulation of blood pressure due to its hypotensive, diuretic, and natriuretic actions. The assay can accurately determine the quantity of ANP in 0.05 ml to 0.10 ml of unextracted normal plasma. The assay has a minimum sensitivity of 2 pg ANP per tube.

Reader Service No. 122.

- The Ribi Adjuvant System (RAS) from Ribi ImmunoChem Research, Inc. is an easy-to-prepare adjuvant system intended as an alternative to the classical Freund's water-in-oil emulsion. RAS is a stable oil-in-water emulsion. Adjuvanticity is improved when the adjuvant is incorporated together with the antigen into the oil phase, rather than incorporating the antigen into the aqueous phase. The mixture is then dispersed in a large volume of saline containing microscopic oil droplets. This makes it possible to have a reduced oil concentration (2% instead of 50%), thereby reducing emulsion viscosity. In RAS the tubercle bacilli contained in complete Freund's adjuvant are replaced by trehalose dimycolate, a highly purified adjuvant. Monophosphoryl lipid A (detoxified endotoxin), *Salmonella typhimurium* mitogen or cell wall skeleton can be added to further enhance immune response.

Reader Service No. 123.

- Seven monoclonal antibodies specific for cytokeratin intermediate filament proteins are now available from Amersham International. This range includes antibodies for use with formalin fixed paraffin embedded material, having an expected application in the sub-typing of normal and pathological epithelia. Also included are antibodies recommended for use with unfixed frozen material. All these antibodies can be used for immunofluorescence staining of cultured cells and 'Western' blot analysis of cytokeratin proteins.

Reader Service No. 139.

- Gibco/BRL's range of products for immunology consists of monoclonal antibodies, hybridoma screening kits and reagents for immunoprecipitation and purification of immunoglobulins. Products for immunodetection based on streptavidin-biotin technology include biotinylated secondary antibodies, streptavidin conjugates and enzyme substrates designed for use in the detection of target antigens by various immunoassay procedures such as ELISA, immunoblotting and immunocytochemistry. For immunoprecipitation BRL offers streptavidin agarose, protein A agarose and immunoprecipitin. For further information on all Gibco/BRL streptavidin-related products use the reader service code to send for a brochure illustrating the applications of streptavidin in the detection of protein, cells and DNA.

Reader Service No. 124.

• Bio-Rad has produced a new publication giving details of a range of high-quality instruments, systems and reagents for purifying and characterizing monoclonal antibodies. The booklet provides information on techniques for purifying antigen by gel electrophoresis and chromatography, screening hybridomas and characterizing hybridoma-secreted immunoglobulins. In addition, there are descriptions of HPLC purification systems and applications for monoclonal antibodies. *Reader Service No. 125.*

• ICN Biochemicals announces the availability of Lipid A, a biologically and immunologically active toxic component of lipopolysaccharides. It has been isolated separately from the serologically specific polysaccharide fraction. Lipid A is known to elicit endotoxic responses, and the availability of this purified form should make it easier to study those responses without the problem of interference from the polysaccharide fraction of lipopolysaccharide. *Reader Service No. 126.*

• Eight new affinity purified antibodies (APAs) to human IgG (Fc), human IgG (Fab) and their corresponding conjugates (FITC, peroxidase, alkaline phosphatase) have been added to Bio-Yeda's range. Products in this series comprise APA to mouse IgG (Fc) and mouse IgG (Fab) and conjugates, and APA to rabbit IgG (H+L). Bio-Yeda APAs and conjugates are purified from goat serum by affinity chromatography with immobilized antigens. At least 95 per cent of the eluted antibodies are active as determined by use of quantitative immunoprecipitation analyses. Because of minimal interspecies crossreactivities to mouse and rat serum proteins, the APAs to human Fab and Fc and their respective conjugates may be used for screening of human monoclonal antibodies produced by hybridoma cells grown *in vitro*. The affinity purified anti-human Fab reagents offer the advantage of increased sensitivity for all the possible human immunoglobulin isotypes (IgG all subclasses, IgA, IgM, IgD, IgE) in biological fluids of tissues in normal or pathological situations. APAs to human Fc and respective conjugates detect specifically human IgG with its four subclasses IgG₁, IgG₂, IgG₃ and IgG₄. The conjugates of these anti-human immunoglobulin reagents are very effective as secondary reagents in ELISA, immunoblotting and immunohistology. *Reader Service No. 127.*

• RIA (UK) Ltd has introduced the Histoclone range of immunohistochemistry kits, based on a series of eight monoclonal antibodies. Designed for the specific identification of human antigens, the kits use routinely fixed, paraffin-embedded tissue samples, or frozen tissue sections. The monoclonal antibodies have been raised against eight diagnostically crucial antigens: luteinizing hormone, human chorionic gonadotropin, growth hor-

mone, thyroid stimulating hormone, prolactin, alpha-fetoprotein, carcinoembryonic antigen and human B cells/Ia-like. *Reader Service No. 128.*

• New from Biotest-Serum-Institut GmbH is a monoclonal antibody against interleukin-2 receptor (Clonab IL-2 R) which detects human activated proliferating T cells. Extremely high purity allows 10 functional tests and more than 1,000 immunofluorescence tests from 1 ml of lyophilized preparation (1 µg of highly purified antibody). *Reader Service No. 129.*

• Becton Dickinson has introduced a new second-antibody for immunofluorescence staining. The monoclonal rat anti-mouse kappa is conjugated to phycoerythrin as a bright second-step reagent for indirect analysis. It is significantly brighter than a FITC conjugated second-step, when used in flow cytometry or microscopy. Anti-mouse kappa phycoerythrin detects low-density antigens where resolution between stained and unstained cells is critical. *Reader Service No. 130.*

• Costar's new Transtar-96 is a portable 96-well liquid handling system with an adjustable volume range of 25–200 µl. The system is completely autoclavable for use in a sterile environment. The heart of the system is a polystyrene radiation-sterilized disposable unit with 96 pipette tips moulded in one piece, covered by a medical-grade silicone rubber membrane. Applications include changing cell culture medium and liquid transfer for assays, for instance in screening of monoclonal antibodies. The 96-well microplate with low evaporation lid minimizes overall evaporation and eliminates "edge effect". *Reader Service No. 131.*

• Metachem Diagnostics' range of specialist antisera for immunocytochemistry now includes antisera to corticotropin releasing factor, growth hormone releasing factor, PHM-27, calcitonin gene related peptide, neuropeptide Y, galanin and peptide YY. The antisera are lyophilized and sealed under nitrogen. After reconstitution the antiserum is ready for use for either PAP or FITC techniques. Supplied with each antiserum is a data sheet with recommended immunohistochemical protocol. *Reader Service No. 132.*

• Cambridge Research Biochemicals have extended their range of peptide antisera to include polyclonal antisera raised to human neurone specific enolase for radioimmunoassay. Pure human neurone-specific enolase is also available. *Reader Service No. 133.*

• The *ras* family of oncogenes are the oncogenes most commonly found in human tumours and have been associated with bladder, lung and colon carcinomas, and various leukaemias. Because the proteins produced from *ras* oncogenes have molecular weights of 21,000, they are referred to as "p21" proteins. The Cetus *ras*

oncogene product monoclonal antibody can be used for experiments designed to explore *ras* oncogene expression in tumours and mutant-specific polyclonal antibodies can be used to detect *ras* mutations, resulting in single amino acid changes in p21. These single amino acid mutations activate the *ras* proto-oncogene to the oncogene form. Rabbit polyclonal antibodies capable of detecting oncogene forms of p21 with serine, valine, arginine or aspartate at position 12, and anti-p21 pan-reactive monoclonal and polyclonal antibodies were introduced at the *Nature* Conference in San Francisco last month. *Reader Service No. 134.*

• The latest Immunochemical Products catalogue from Zymed includes a long list of affinity purified antibodies to mouse, human and rat immunoglobulins. Each antibody is characterized according to its reactivity to a specific class of immunoglobulins and cross-species reactivity. The streptavidin/biotin system is of particular interest to those who need to enhance the signal for enzyme immunoassay, histochemical and cytochemical staining whilst retaining negligible background. Monoclonal antibodies are also listed as are recommended methods used in immunology. *Reader Service No. 135.*

• Organon Teknika has introduced what is claimed to be the first non-isotopic immunoassay for FSH and LH in serum or plasma. Both tests, FSH-Nostika and LH-Nostika, are based on the sandwich principle for the detection of physiological values of the respective hormones. Incubations are at room temperature, and results are available on the same day (total incubation of 4½ hours). The system gives good reproducibility and has the same sensitivity as radioimmunoassays. *Reader Service No. 136.*

• The new disposable affinity columns from Biological Consultancy Services (BCS) make it possible to purify mouse or human IgG, IgM, IgA from serum, tissue culture fluid or ascites in one step. Each column can be used several hundred times: thus they will purify immunoglobulin from as little as 1 ml fluid or as much as several litres. Immunoglobulin can be purified to >99% purity in one step. There is no need to concentrate the starting fluid or to partially purify by ion-exchange or molecular size chromatography. *Reader Service No. 137.*

• Labsystems' Uniskan range of photometers includes the new Uniskan II, designed for use in absorbent and turbidimetric measurements typical of enzyme immunoassay and microbiological work. Uniskan's optics are based on vertical photometry which eliminates the adverse effects of layering, evaporation and other sources of error. The life of the light source has been extended to a quarter of a million measurements by means of pulsing the beam. *Reader Service No. 138.*